

# From Monologue to Dialogue?

## Cohesive Interaction in Personal Weblogs

Inauguraldissertation  
zur Erlangung des Doktorgrades der Philosophie  
(Fach: Englische Sprachwissenschaft)  
an der Universität Augsburg  
Philologisch-Historische Fakultät

vorgelegt von Christian R. Hoffmann

1. Gutachter: Prof. Dr. Wolfram Bublitz  
2. Gutachter: PD Dr. Christoph Schubert  
Disputationstermin: 10.03.2010, 15.00 Uhr

### Eidesstattliche Versicherung

Hiermit versichere ich, Christian Rainer Hoffmann, geboren am 27.05.1979, dass die vorliegende Dissertation mit dem Thema:

*From Monologue to Dialogue? Cohesive Interaction in Personal Weblogs*

selbstständig und ausschließlich mit den im Literaturverzeichnis angegebenen Quellen und Hilfsmitteln angefertigt wurde.

Augsburg, den 07.01.2010

---

für Berit

Zwar ist es leicht, doch ist das Leichte schwer;  
Es liegt schon da, doch um es zu erlangen,  
Das ist die Kunst, wer weiß es anzufangen?

*Faust II, 1. Akt*

So eine Arbeit wird eigentlich nie fertig,  
man muß sie für fertig erklären,  
wenn man nach Zeit und Umständen  
das Mögliche getan hat.

*Goethe, Italienische Reise, 1787*

## **Danksagung**

Mein besonderer Dank gilt zunächst meinem Doktorvater, Prof. Dr. Wolfram Bublitz. Er stand mir in jeder Phase der Entstehung dieser Dissertation immer als Ansprechpartner in fachlichen, methodischen, aber auch terminlichen Nöten zur Seite.

Ein großer Dank gilt daher ihm, sowie auch meiner Kollegin Jenny Arendholz und meinem Kollegen Volker Eisenlauer. Die Zusammenarbeit mit ihnen war ein Meilenstein bei der Erstellung dieser Arbeit. Mit ihrem fundierten Fachwissen haben sie mir zahlreiche wertvolle Anregungen für meine wissenschaftliche Arbeit gegeben, die mir sehr geholfen haben. Es besteht kein Zweifel, dass diese Arbeit ohne ihr Wissen, ihre Ideen und ihre konstruktive Kritik niemals soweit gekommen wäre.

Darüber hinaus möchte ich mich bei Siaohui Kok, Angelika Behn und Olena Vorontsova bedanken, die mich gerade in der Endphase beim Korrekturlesen freundlich unterstützt haben. Besonders Siaohuis muttersprachliche Kompetenz war mir vor allem in vielen sprachlichen Feinheiten eine große Hilfe bei der Überarbeitung der Texte. Für sämtliche Fehler und Versäumnisse, die noch im Text verblieben sind, bleiben natürlich alleine ich selbst, ein gewisser Mangel an Schlaf und MS Word verantwortlich.

Einen Ort der Ruhe und des Zuspruches fand ich innerhalb der Universität während der intensiven Arbeitsphasen auch immer wieder im Sekretariat. Dafür, sowie für ihre hilfreiche administrative Unterstützung bin ich Monika Martens und Gudrun Nelle zu Dank verpflichtet.

Nicht zuletzt möchte ich auch den vielen Bloggerinnen und Bloggern bedanken, die mir so uneigennützig ihr Datenmaterial für diese Erstellung dieser Arbeit zur Verfügung gestellt haben.

Schließlich möchte ich mich sehr herzlich bei meinen Eltern bedanken, ohne die mein Studium und diese Doktorarbeit niemals möglich geworden wären. Euch, Susanne, Michael und Oma sowie allen Haugs danke ich für die fortwährende Unterstützung. Ein besonderer Dank geht natürlich an meine Frau Berit, die mich stets bestärkt hat, wenn ich an mir gezweifelt habe.

In Gedanken geht auch ein großer Dank an zwei Großväter, die wir während der Entstehung dieser Arbeit verloren haben, und mit denen wir ihr Gelingen gerne gefeiert hätten.



## Table of Contents

<b>1. The Objective</b>	<b>5</b>
1.1. Introduction	5
1.2. The Weblog between Monologue and Dialogue	6
1.3. Text and Discourse	8
1.4. Discourse Analysis: Two Vantage Points	10
1.5. Cohesion and Coherence	11
1.6. Aims and Outline of the Study	12
 <b>2. The Object</b>	 <b>14</b>
2.1. Defining the Weblog	14
2.2. The Formal Composition of Weblogs	20
2.2.1. The Upper Panel	21
2.2.2. The Side Panels	23
2.2.3. The Lower Panel	24
2.2.4. The Entries	25
2.2.5. The Comments	26
 <b>3. The Genre</b>	 <b>28</b>
3.1. The Naturalization of the Blogosphere	28
3.2. Diary, Journal or Weblog – A Path Toward Generic Attribution	32
3.3. Personal Weblogs as a Meta-Genre	38
3.4. Weblogs as Hypertext	43
3.5. Three Perspectives on Hypertext	44
3.6. Across Media – From Analogue to Digital Hypertext	50
3.7. Across the Mind – Hypertext and Cognition	54
3.8. Across Space – Knowledge, Discourse and Participation	59
 <b>4. The Texture: Verbal Cohesion</b>	 <b>69</b>
4.1. A Framework for Verbal Cohesion in Weblogs	69
4.2. The Scope of Cohesive Relations	70
4.3. Grammatical Cohesion	72
4.3.1. Reference	73
4.3.2. Substitution and Ellipsis	80
4.3.3. Conjunction	84
4.4. Lexical Cohesion	85
4.4.1. Repetition	87
4.4.2. Equivalence	88
4.4.3. Superordinate	90
4.4.4. Co-Hyponymy	91
4.4.5. Antonymy	91
4.4.6. Collocation	92

<b>5. The Corpus</b>	96
5.1. The Augsburg Weblog Corpus (AWC)	96
5.2. The Segmentation of Data	99
5.3. Manual Analysis and Final Evaluation	103
5.4. Preliminary Theoretical Reflections	106
<b>6. The Analysis I: Grammatical Cohesion</b>	109
6.1. Reference in Weblog Entries	109
6.2. Reference in Weblog Comments	126
6.3. Substitution in Weblog Entries and Comments	143
6.4. Conjunction in Weblog Entries and Comments	148
6.5. Ellipsis in Weblog Entries and Comments	158
6.6. Some Preliminary Results	162
<b>7. The Analysis II: Lexical Cohesion</b>	164
7.1. Lexical Cohesion in Weblog Entries	165
7.2. Lexical Cohesion in Weblog Comments	181
7.3. Some Preliminary Results	190
<b>8. The Interaction: Knowledge and Cohesion</b>	192
8.1. From Collocation to Cognition	192
8.2. Serial Knowledge and Episodic Memory	206
8.3. A Brief Analysis of Weblog Interaction	210
<b>9. The Results</b>	216
9.1. Cohesive Interaction Revisited	216
9.2. Monologue or Dialogue? – Positioning Weblog Discourse	217
9.3. Communicative Conditions in Personal Weblogs	222
9.4. Limitations of the Study and Future Research	225
9.5. Concluding Remarks	227
<b>10. References</b>	228
<b>11. Bibliography</b>	239
<b>12. Appendix</b>	240



## Typographic Conventions

---

### *Bold type*

for emphasis in citations or anaphora/cataphora in text examples

### *Italics*

for technical terms when first introduced

### *Square brackets + capital letters*

for cognitive frames or concepts and ellipsis in examples

### *Underlined items*

for co-referents (e.g. antecedents) in text examples

## List of Abbreviations for AWC Weblogs

---

<i><b>Weblog Title</b></i>	<i><b>Abbreviation</b></i>
All Adither	AA
Back Home Again	BHA
Beetle's Memories and Ravings	BMR
Blog of a Good Time	BOGT
Five Blondes	FB
Mushy's Moochings	MM
NonSoccerMom	NSM
Oh my word!	OMW
Shelli's Sentiments	ShSe
Whatever I Think	WIT

## The Objective

---

The Internet is the beginning of a new stage in the evolution of the written language [...]. It is exciting to be in at the outset of it. But it will be even more exciting to see what happens when the medium evolves a more routine spoken dimension, and when the results of streaming different modalities foster fresh forms of expression [...].  
(Crystal, David 2005:3f.)

### 1.1 Introduction

Since the turn of the last century, the Internet has evolved into one the fastest growing resources for linguistic research. Originally intended for the quick retrieval and exchange of scientific information, the Internet has since long become a platform for manifold communicative practices. Its rapid socio-technological evolution during the last two decades has led to the emergence of novel forms of communication, e.g. Internet chats, wikis, weblogs and social network sites (SNS). One of these new writing spaces has proven particularly adaptable to the Internet's ongoing transformation, namely weblogs. It is weblogs' chameleonic nature which is primarily responsible for their growing popularity among Internet users worldwide. In 2007, the Technorati state of the blogosphere report already tracked an astonishing number of 70 million weblogs, with an additional "120,000 new weblogs being created worldwide each day".<sup>1</sup> One year later, two independent studies (*comscore Media Matrix* and *Universal McCann*) already captured an additional increase of seven million new weblogs.

This rapid expansion of the *blogosphere* has certainly changed the formal and functional appeal of weblogs. It has contributed to their technological and communicative emancipation from websites and has spawned a plethora of weblog genres.<sup>2</sup> Their topics and purposes are diverse: users can contact, build up and maintain friendships, store and exchange past experience or use weblogs as a professional tool for various educational, organisational or commercial purposes. Differences between these genres equally emerge in different writing styles. Hence, weblogs range from erratic scribbling to bloggers' careful reflections on complex

---

<sup>1</sup> (<http://Internet.sifry.com/alerts/archives/000493.html>, 20/04/09).

<sup>2</sup> The blogosphere is the interconnected community of weblog authors and readers on the Internet (cf. [Internet.le.ac.uk/cc/glossary/ccglb.html](http://Internet.le.ac.uk/cc/glossary/ccglb.html)).

social, economical or political affairs.<sup>3</sup> This abundance of communicative shapes and genres is still largely unexplored.

## 1.2 The Weblog between Monologue and Dialogue

To this date, linguistic research on weblogs is scarce although a great deal of weblog research is currently underway as I am writing these pages. Following Androutsopoulos (2008), previous research on computer-mediated communication (CMC) can be split into two main phases or waves. At the turn of the century, a first wave primarily focussed on the general nature of “Internet language” as opposed to traditional spoken and written language. Related studies usually retrieved data randomly from the Internet, “detached from their discursive and social contexts” (Androutsopoulos 2008: not paginated). In contrast, a second wave of CMC research has only recently started to centre on the discursive and socio-pragmatic dimension of individual text genres on the Internet. With respect to weblog communication, this second “wave” connects to upcoming work on the discursive and pragmatic dimension of weblog genres. It is argued that the corpus-based description and context-sensitive evaluation of discourse yields more accurate results about the way people communicate via computers.

The lack of corpus-based studies on weblog genres gives rise to great speculation on the nature of weblog discourse both in science and in mainstream media. Weblogs are described inconsistently and characterized either as written conversations or spoken monologues, written dialogues or written monologues. To this end, it is interesting to note that most CMC research seems to highlight an alleged conversational quality of weblogs. It is not only believed that weblogs comprise a certain degree of *konzeptionelle Mündlichkeit* (conceptual orality, Koch & Oesterreicher 1995:19-21) but also that weblogs *per se* bring about an “ideal speech situation” (Wijnia 2005:38). However, few actually indicate how this new function of weblogs actually manifests itself on a formal or discursive plane.

It is, of course, true that weblog software enables Internet users to respond to weblog posts (entries). They can upload their own contributions to the weblog and

---

<sup>3</sup> In fact, current reports in the mass media usually focus on the opinions of few popular bloggers which, apart from setting the agenda in the blogosphere, are equally starting to assert discursive power on cultural, economical and political matters in the “offline” world.

therefore engage in a quasi-dialogical interaction. It is this potential for dialogicity which is usually held responsible for an inherent dialogicity of weblog discourse:<sup>4</sup>

It could be argued that blogs combine both the monologue and the dialogue in a space-bound electronic environment. They are simultaneously self-reflective thoughts presented publically, and continuous conversations.  
(Nilsson 2003:31)

The degree to which weblogs are dialogic has been vigorously debated in the circles of both bloggers and researchers. Although technologically, weblogs appear to bridge a gap between website's monologue and Internet chat's polyphony, it remains unclear whether weblog users embrace the interactive potential.<sup>5</sup> Evidently, the "conversational nature of blogging has yet to be empirically investigated" on a discursive plane (Herring 2005: no pagination). In this light, this study aims to find out the actual extent of discursive collaboration between interlocutors in weblogs. By investigating the interactive distribution of cohesion in and across weblogs, the distinct cohesive profile of weblogs will be revealed and compared to the profiles of written monologues and spoken dialogues. It is argued that this comparison based on corpus-derived empirical sets of data promises more valuable insights into the discursive realm of weblogs than the mere quantification of syntactic or single lexical forms. In this light, I shall postulate five research questions which I seek to answer in the course of this study:

1. Do weblog entries and comments make different use of cohesive means?
2. How exactly do weblog entries and comments differ with respect to their variation and distribution of grammatical means of cohesion?
3. How exactly do weblog entries and comments differ with respect to their variation and distribution of lexical means of cohesion?
4. Does the cohesive profile of weblog discourse resemble prototypical spoken dialogues, prototypical written monologues or both?
5. How interactive is weblog discourse in and across entries and comments?

The quantification of individual formal means promises little insight into the way weblog authors and users engage in their discursive practice. In contrast, a cross-

---

<sup>4</sup> Voiskounsky (1997), for instance, speaks of different conversational settings in computer-mediated communication; a phenomenon he calls *telelogues*.

<sup>5</sup> Precisely this communicative "double-play" defines the focal point of current British and U.S. American corporate media coverage of weblogs (<http://Internet.bivingsreport.com/2007/analyzing-the-websites-of-american-magazines/>).

generic investigation of cohesive means can be expected to yield better results on the nature and amount of discourse interaction in weblogs. Before I will present the research design of this study, I will now briefly define some of the most central terms and concepts recurrently applied throughout the thesis. The following will thus put the study on a proper theoretical footing and provide the necessary theoretical background for the ensuing empirical research.

### **1.3 Text and Discourse**

*Text* and *discourse* are two of the most frequently used and hotly debated terms in discourse analysis. It seems sensible to explain what I shall wish to mean by *text* and *discourse* in the course of this study. The way I perceive of both terms is closely informed by the way I define meaning. Following Bublitz (2000, 2006), I advocate a constructivist and hermeneutic stance on discourse meaning. Meaning is not physically present in a given text but rather ascribed to texts by interlocutors in ongoing interaction. It emerges rather naturally and unconsciously in any new act of interpretation, and it is only when one fails to understand that the constructive character of meaning becomes apparent. Meaning-making is both a mental and social process. Meaning arises as a temporary by-product of our active hermeneutic engagement with a semiotic artefact. This can be an auditory signal (e.g. *voice*) or a visual perception (e.g. *document, picture, film*). In any case, we engage with a material entity of some sort. I will call this substance a *text*. Texts can be in auditory (spoken texts) or visual (written texts, videos, pictures) form. They may be large (an advertising banner hanging from the outer wall of a skyscraper) or small (a sticker on a tin can).

When we engage with texts, we put them into perspective in a given situation or context. We activate our personal knowledge to ascribe meaning to the text. Through this active engagement, the text turns into discourse which can be defined as *text-as-interpreted* (cf. Widdowson 2008). A text is physically given, which means it might be read by different people at different times in different situations. Hence texts have meaning potential. Discourse, in contrast, is created in every new individual act of interpretation. If we find out something new about the text, the discourse changes because our interpretation of the text has been altered. Discourse therefore relates to the individual. It is temporary and bound to a certain context of use. In short, texts exist as material works, while discourse need to be created. If we

turn text into discourse, we can only arrive at a fragmentary, temporary understanding of a text. The meanings we project in ongoing discourse are necessarily transient since we establish them on the fly. They are never final. Authors of texts can, of course, try to indicate discourse meaning in their speaking and writing, and usually they are quite successful in doing so. In fact, hearers or readers are expected to use these textual indications to accord a specific discourse meaning to the text. Discourse thus reflects the constructive process of interpreting text-based meanings. Nonetheless, the process is entirely driven by the (inter-)actions we take to achieve these goals. What remains is the negotiation of semiotic traces of meaning induced by the speaker/author and interpreted by the hearer/reader.

<i>Text</i>	<i>Discourse</i>
closed	Open
fixed	Dynamic
context-free	context-bound
completed	Negotiable
post-festum	in actu

Table 1: Main Properties of *text* and *discourse*

From this perspective, discourse is characterized by a profound interpersonal dimension, for it is only through social conduct that we may gather discourse meaning.<sup>6</sup> We can see now that discourse is the intersection between meaning as a mental and social concept. As a result, one of the primary goals of discourse analysis has been the description of textual means which gear the practice of meaning negotiation in discourse. These means, which enhance our ability to construe meaningful discourse, are called *cohesive means*. The investigation of these means and their distribution in weblogs will take centre stage in this study, not least because they largely catalyze and organize discourse interaction. If we analyse cohesive relations in weblogs, we interpret cohesion as a discursive tool or instrument towards the collaborative genesis of monologic or dialogic discourse.

---

<sup>6</sup> Since the process of interpretation requires readers to engage cognitively with the creative product of a different person, reading a book can be regarded as a social activity.

#### 1.4 Discourse Analysis: Two Vantage Points

We can describe cohesive means from two perspectives: an author-centred and a reader-centred perspective. The author-centred perspective gives priority to the creative acts of text production, i.e. to the role of speakers or authors. It revolves around the observation that text creators create purposeful texts for a particular target audience. In order to achieve their communicative aims, authors apply specific formal means which are meant to guide recipients through the texts. These means trigger specific interpretations of texts. As such, the author-centred perspective sheds light upon the principles authors follow when they choose and arrange meaningful signs.<sup>7</sup>

In contrast, the reader-centred perspective does not centre on the production of text but rather on its interpretation. Accordingly, texts have only meaning potential. However, it is in the act of interpretation that meaning is actually created. Therefore, the act of creation which is focussed now shifts from textual production to discourse creation. If we follow this line of thought, we can concur that meaning need not be automatically inferred from cohesive cues in a given text. Although cohesive cues may be informative to the interpretation of a text, they need not be necessary. Previous research has shown that sometimes a text may be fully comprehensible to recipients without cohesive means, exclusively through its *context* (cf. Bublitz 1994:219).<sup>8</sup> Context may not only be composed of elements of the discursive environment but may equally relate to the minds of individual interlocutors. For instance, we may need to have specific knowledge about a state of affairs to understand a stretch of discourse. We connect what we know with what we read: our individual knowledge provides the context for our individual interpretation. These contextual cues, either of situational or cognitive sorts, are called *exophoric* because they are external to the ongoing discourse.

In some way, the notions of *text* and *discourse* thus reflect two main theoretical positions advocated in discourse analysis. In other words, we may approach text-based meaning from an authorial point of view, which describes discourse as mode of production but we may just as well define text as an interpretative notion and talk about discourse instead. It is conceivable that these two positions can be reconciled quite simply because communication is a dual concept

---

<sup>7</sup> I shall henceforth use the generic pronoun “he” to address both male and female referents alike.

<sup>8</sup> When (non-)textual environment becomes viable for the interpretation of discourse, it becomes *context*.



which involves both text production and discourse interpretation. One perspective presupposes the other. There is a middle ground between both positions. I contend that meaning-making is a negotiable practice which is mutually operated by producers and recipients. The semiotic currency through which they *mainly* organize their discursive moves is cohesion. Authors regularly apply large amounts of cohesive means to indicate meanings in their texts and readers draw on them in an effort to create coherent, meaningful discourse. In interactive settings, readers can even turn into authors and add cohesive means of their own to a given text. In this case, the cohesive texture becomes interactive (or dialogical). We can conclude that cohesion seems to define a centre between author- and reader-centred perspectives. It is this reconciliatory position which seems most beneficial for the purposes of this study.

### 1.5 Cohesion and Coherence

It seems obvious that an author-centred perspective of discourse is biased toward the cohesive means which individual authors deliberately include in their texts. It thus mainly endorses cohesion as a text-centred notion since the cohesive means are given as material, formal cues inside the texts. Author-centred views on discourse often assume that an adequate application and distribution of cohesive means renders texts *coherent*, while an insufficient use of cohesive means will lead to cases of disturbed coherence.

I have just explained that discourse may perfectly well be *coherent* without necessarily being cohesive at the same time. Hence readers interpret discourse not by itself but in *context*. What we call context here either refers contiguous textual information, i.e. the *co-text*, or to external information provided by the situation in which the discourse takes place or filled in by interlocutors' knowledge about the state of affairs. Therefore, it is not only the discourse itself which becomes relevant for making texts coherent but the identification of a relation between discourse and context.

It follows that whereas cohesion is fixed in textual form, coherence is a mental process or product. Discourse acquires meaning only to the extent that one is able to integrate *what is said* with *what is meant* in a certain place and at a specific time (situation) by connecting it to what is known (knowledge). In its contextual dimension, coherence is individual, dynamic. It is an inherently unstable concept. It can be

defined as the temporary result of discursive projections between the three spheres of discourse, situation and mind.

Although cohesion may thus be insufficient to explain how readers turn texts into coherent discourse, we may still contend that cohesion usually contributes considerably to the construal of coherence. In fact, in two-sided forms of communication, speakers/authors and hearers/readers regularly change their roles in interaction. In face-to-face conversation, for instance, interlocutors switch back and forth between active text creation (speaker role) and active discourse construal (hearer role). As weblogs equally allow for two-sided forms of communication (through weblog comments), we shall hold that the construction of cohesion and coherence is a dynamic and collaborative undertaking. Cohesion thus drops its static meaning and becomes a formal nexus of interaction between meaning indication and interpretation. A more elaborate understanding of discourse cohesion can thus be expected to yield a more comprehensive understanding on how we collaborate toward construing meaningful discourse.

## **1.6 Aims and Outline of the Study**

This study aims to examine the way in which weblog authors and users make use of grammatical and lexical means to construe meaningful discourse. A comparative analysis of the variation and distribution of cohesion in weblog discourse will provide new insights into the discursive collaboration of interlocutors in weblogs. It reveals the degree to which cohesion in the most pervasive genre of weblogs, personal weblogs, resembles prototypical written monologues and/or spoken dialogues.

Methodologically, the study comprises the following chapters. Chapter two and three introduce the concept of weblogs and describe it in its formal, functional, historical dimension. Chapter three additionally presents a viable generic classification for weblog genres and explains the most essential ramifications of the hypertext paradigm for the analysis of weblog discourse. Chapter four then develops a suitable framework for the quantitative and qualitative analysis of weblog cohesion, drawing on previous linguistic research on cohesion and coherence, such as Halliday & Hasan 1976, Martin 1992, Tanskanen 2006, Schubert 2008.

In chapter five, the methodological stages which have led to the compilation of the self-compiled weblog corpus will be introduced. The chapter primarily centres on the theoretical complications and methodological implications of compiling weblog

language data for discourse analysis. Chapter six concurrently applies the cohesive framework developed in the previous chapter to the corpus developed in chapter six. First results on the variation and distribution of grammatical cohesion in and across weblogs will be presented and interpreted here. Chapter seven reveals additional results from the analysis of lexical means of cohesion in weblog entries and comments. These results shall be compared to compatible cohesive profiles of spoken dialogues (two-party and three-party conversations) and written monologues (academic articles) as proposed in a recent study by Tanskanen (2006). The comparison of cohesive profiles across text genres will allow a more precise interpretation of the monologic or dialogic quality of personal weblogs.

In chapter nine, the study will close with an exhaustive examination of the interrelation between weblog cohesion, cognitive frames and participation roles. I shall illustrate how cohesive collocation gives rise to the negotiation of common ground between interlocutors in weblog discourse. In this last chapter, I aim to prove that cohesion and coherence are deeply entrenched hermeneutic concepts which cannot be easily disconnected. Therefore, this study hopes to be part of a pioneering chapter of linguistic work on weblog genres. Hopefully, it will be a springboard for future studies which illuminate the exciting discursive and socio-pragmatic aspects of an intriguing form of communication.

## The Object

---

A true-born child of the computer medium, the weblog may have its roots in the research journal, the ship's log, the private diary and the newspapers, all at the same time. But like a mongrel hunting the dark alleys of the digital city, the weblog is nothing if not adaptive and unique at the same time. No fancy thoroughbred this [...] but a bastard child of all personal writing, breeding wildly as it meets others of its ilk online.  
(Mortensen 2008)

### 2.1 Defining the Weblog

Defining the general shape and usage of weblogs is a highly delicate (some say impossible) task, given the metamorphic character of this phenomenon as well as its vast extension on the Internet. Many conceptual pitfalls need to be mastered in order to arrive at a viable verbalization which is capable of encapsulating the multifarious shapes of weblogs. In order to cope with these descriptive obstacles, I shall begin with delineating the object of this study in search of elementary patterns in a number of influential weblog definitions. These patterns or characteristics describe different viewpoints which have claimed descriptive primacy in certain chronological periods of the generic evolution of weblogs. By comparing these interrelated concepts and stages, I hope to be able to slowly excavate weblogs' central and peripheral criteria.

Writers of both mundane and scholarly provenance have so far taken on the intricate task of describing what weblogs actually are, and sometimes rather what they should be. In fact, there exists an amazing plurality of pre- and descriptive attempts to define weblogs all of which pertain to individual selections of compositional parts, text structures, themes or functions. It is interesting to note that the definition of weblog essentials seems to have altered from the late 1990's to the present day, and this "shifting" momentum affirms the persistence of some as well as the slow demise of other weblog features over time. A set of scientifically acclaimed definitions should by comparison enable me to disclose some primary conceptual rifts which are epitomized by these definitional changes, sketching out the main chronological stages of weblogs' structural and functional evolution. In other words, I take single weblog definitions to reify descriptive patterns representative of a particular phase in what Heyd (2009:245) coins the "genrefication" of weblogs. Toward capturing these stages, the following scientific resources were consulted to surface specific weblog properties:

- Barger 1997 (BAR)
- The Oxford English Dictionary 2003 (OED)
- Blood 2003 (BLO)
- Herring et al 2004 (HER)
- Miller & Shepherd 2004 (MIL & SHE)
- Walker 2005 (WAL)
- The Merriam-Webster Online Dictionary 2008 (MWO)

The term “web-log” was allegedly first coined by John Barger who in 1997 used it to refer to web pages which exhibited collections of annotated hyperlinks on a certain topic.<sup>9</sup> At that time, weblogs were no more than regular websites which made extensive use of hyperlinks to update like-minded Internet users about the latest news. Only a couple of years later did weblogs already assumed their - now so popular - diary shape, prototypically reflected in the “weblog” entry of the current online edition of the Merriam-Webster dictionary (henceforth: MWO). The entry, dating back to 1999, informs us that a weblog is

a website that contains an online personal journal with reflections, comments, and often hyperlinks provided by the writer.  
(<http://Internet.merriam-webster.com/dictionary/weblog>)

In a very similar manner, the Oxford English Dictionary (henceforth: OED) presents its definition of the weblog in the following:

A frequently updated website consisting of personal observations, excerpts from other sources, etc. typically run by a single person, and usually with hyperlinks to other sites; an online journal or diary.  
(Oxford English Dictionary 2003)

Note that both descriptions above acknowledge the fact that weblogs represent a new evolutionary phase of websites, i.e. they are built with similar computer codes (HTML/XML). Also both descriptions emphasise the importance of computer technology for blogging such as their specific use of hyperlinks, possibly to contrast weblogs with traditional diaries. They equally agree that weblogs are typically written by one individual only rather than by a group of co-writers. On this basis, we may thus deduce a first subset of weblog properties, namely [origin] = websites, [technology] = hyperlinks, [genre] = personal journal.

Around the same time, in 2003, Rebecca Blood advanced an influential weblog definition which is usually referred to in current weblog research (cf. Herring 2004, Walker 2007). Although Blood’s definition resembles the two aforementioned

---

<sup>9</sup> See Blood (2003:7)

definitions, it also excludes some of their previous properties. More precisely, Blood (2003) reduces weblogs simply to “websites which are up-dated frequently, with new material posted at the top of the page” (Blood 2003:IX). This conception of weblogs differs with regard to two crucial aspects from the two dictionary-based ones. On the one hand, Blood does not assign specific genre status to weblogs which has been done in both the WMO and OED definitions. On the other hand, Blood shifts the attention from generic attribution to particular compositional traits of weblogs, for the first time defining a typical entry arrangement in reverse chronological order; a feature which was fully neglected in the MWO and OED texts.<sup>10</sup>

Three years later, literary scholar, Jill Walker advanced an elaborate 500 word weblog definition in the *Routledge Encyclopaedia of Narrative*. What is interesting about Walker’s definition is the process of its creation. Posting a first draft of the definition on her own weblog ([www.jill/txt.net](http://www.jill/txt.net)), Walker encouraged users and fellow bloggers to discuss the definition in a related comment section. Within the bounds of this interactive approach, Walker managed to reflect a blend of her personal experience of weblogs with the collaborative evaluations and comments of other bloggers. In a way, Walker therefore comes closest to merging her own experience and introspection with the opinions of bloggers equally which engaged in the blogging process. Through her collaborative attempt to describe weblogs, Walker thus evades a regular tendency in weblog research to describe weblogs “[...] quite distinct from and in fact in conflict with actual practice” (McNeil 2009:144). She defines weblogs in the following way:

A weblog, or blog, is a frequently updated website consisting of dated entries arranged in reverse chronological order so the most recent post appears first [...]. Since anybody with a net connection can publish their own weblog, there is great variety in the quality, content, and ambition of Weblogs, and a weblog may have anywhere from a handful to tens of thousands of daily readers.  
(Walker 2005)

At close sight, Walker’s definition centres on most of the weblog properties already elicited in the previous descriptive attempts, predominantly focussing on the properties of [origin], [technology] and [composition]. In addition to this set of properties, Walker portrays weblogs as elusive and variable textual entities, and she thereby rejects clear-cut generic fixations. As a result, she objects to the stance that

---

<sup>10</sup> Blood’s decision to exclude all reference to technology-related weblog features may also have been informed by the radical growth and frequent usage of Web 2.0 technologies at the time, whose generic status, however, was unclear.

weblogs are primarily representative of the journal or diary genre. Notwithstanding her own objections, Walker (2005) still concedes that weblogs may indeed be “typically [...] published by individuals and their style is personal and informal”. At the same time she predicts a decrease of stylistic and thematic genre markers, tying the weblog to a particular genre. This de-genrefication of weblogs, she claims, was set in motion with the growing access of “free publishing tools toward the turn of the century” (Walker 2005). Walker likewise asserts that the emergence of free-to-use weblog services greatly contributed to boosting their growing popularity by facilitating the generation and use of weblogs for the common user.<sup>11</sup> Moreover, these services would also some years later foster new technological means (usually known as Web 2.0 technology) which would again lead to an increased user engagement in collaborative text production and weblog interaction. In summation, we can thus change our previous list of weblog criteria, from [genre] = personal journal to [interaction] = augmented participation.<sup>12</sup> To this date, the descriptive subset of weblog properties provided by Walker seems to be generally accepted in weblog research. Yet, while Walker’s definition finds wide recognition in the literary sciences, linguist Susan Herring (Herring et al 2004) developed an similar definition of weblogs in linguistics. As part of her work of the university of Indiana’s BROG research project, Herring and her team analysed sets of empirical weblog data, leading up to another repeatedly cited weblog definition. She describes the phenomenon as a “[...] frequently modified web pages in which dated entries are listed in reverse chronological sequence” Herring et al (2004:1). Her definition aligns with two central weblog properties [origin] = web page, [composition] = reverse chronological order. Just as Blood (2003) and Walker (2005), Herring et al (2004) are reluctant to accord distinct generic patterns to the weblog. Rather, she concurs that weblogs are neither unique nor are they reproduced entirely from offline genres. They rather constitute a hybrid genre which “draws from multiple sources, including a range of other Internet genres” (Herring et al 2004). Adhering to the generic hybridity of the blogging phenomenon, Herring et al (2004) then continue to attribute a derivate of journal-like and conversational characteristics to weblogs, thus subtly reinforcing Walker’s notion of generic disparity. Herring et al claim continue claiming that

---

<sup>11</sup> The term *participation* will be understood as two-sided written forms of communication enhanced by Web 2.0 web technologies.

[j]ournal sites, with their lesser interactivity, are closer to standard Web pages than are Weblogs. Community sites are closer to online discussion groups than are individually-maintained Weblogs in their frequency of activity and exchange of messages among multiple participants.  
(Herring et al 2004:10ff.)

The same year, Miller & Shepherd (2004) proposed a weblog definition based on the insights gained from empirical weblog data but also “relying to the greatest extent possible on the perceptions of bloggers themselves” (Miller & Shepherd 2004: no pagination). Anticipating Walker’s interactive approach towards defining weblogs, Miller & Shepherd described weblogs by recourse to empirical data backed up by a series of interviews with weblog users. Their aim was to define Weblogs as a new text genre which emerges from a purposeful combination of genre conventions “imported” from what the authors call “ancestral genres” (Miller & Shepherd 2004). The study’s results revealed the weblog as an unlikely candidate for any sort of stable generic identification. While Weblogs revolve around both structural and thematic traits various classic genres, e.g. the personal journal, the diary, the log book, the newspaper article etc., they also figure in the public sphere of participatory text collaboration:

In the weblog, the potentialities of technology, a set of cultural patterns, rhetorical conventions available in antecedent genres, and the history of the subject have combined to produce a recurrent rhetorical motive that has found a conventional mode of expression.

(Miller & Shepherd 2004: no pagination)

In this context, Miller & Shepherd take us through a set of potential formal, semantic and pragmatic hallmarks of weblogs in an attempt to narrow down their actual generic scope. In the end, however, they end up making only few resilient claims to the function of these components as genre markers. Instead, they inform us that their interviewees managed to demarcate some or all of the following technological and compositional weblog properties as essential blogging standards: [composition] = reverse chronology, frequent updating [technology] = hyperlinks, backtracking, [interaction] = commentary, textual collaboration. Not surprisingly, however, when asked if they could detect a recurrent blogging theme or purpose, no actual claims are made. This fact strengthens Walker’s and Herring’s stance that weblogs’ are hybrids with regards to their topical and functional orientation. Indeed, Miller and Shepherd conclude that a “clear picture” of certain formal and technological



parameters, which would naturally help to set the weblog apart from other “old” media genres, dissolves the moment “bloggers discuss the purpose of the weblog, its function and value as social action involving rhetors and audiences” (Miller & Shepherd 2004: no pagination). In fact, they recognize this circumstance elsewhere even more explicitly when they admit that “it may no longer be accurate to think of the weblog as a single genre” (Miller & Shepherd 2004).

SOURCES	BAR (1997)	MWO (1999)	BLO (2003)	OED (2003)	HER (2004)	SHE/MIL (2004)	WAL (2005)
PROPERTIES							
A [+ origin]	✓	✓	✓	✓	✓		✓
B [+ technology]	✓	✓	✓	✓		✓	✓
C [+ composition]			✓		✓	✓	✓
D [+ genre]		✓		✓			
E [+ interaction]						✓	✓
F [+ private/public]						✓	✓

Table 2: Defining Properties of Weblog Definitions

Table one reveals that weblog definitions have gradually discarded some generic attributions from 2002 onwards. Hence, Herring et al (2004), Miller & Shepherd (2004) and Walker (2005) abstain from using generic parameters but are inclined to incorporate the parameter of interactive participation into their descriptions. Also the technological divide between the private and the public domain seems to have become more substantial in the last years. Whereas early definitions regarded weblogs as texts of a private sphere going public (*privately public*; cf. Lange 2007), later descriptions embraced the notion of weblogs as public fora enriched through personal perspective (*publically private* cf. Lange 2007). In this respect, different functional implications apply each time. While in the first case weblogs are envisaged as channels for self-expression and self-reflection accessible only by a chosen few, they become, in the second case, a tool to discuss collaboratively affairs without the interference of any internal or external censor. In table 1, one can discern a descriptive pattern related to properties A, B & C. Also, I detect a tendency in more recent descriptions to highlight new discursive traits such as weblogs’ various degrees of written interaction and its intermediate positioning between the private and public realm. On this basis, I propose a revised working definition which is largely derived from Walker:

A weblog (or blog) is a frequently updated website consisting of dated entries arranged in reverse chronological order. Weblogs can comprise various audio-visual modes (audioblogs, vlogs, textblogs, etc.) and can be composed by an individual or a group of collaborative writers (bloggers). Likewise, weblogs are readable by an indiscriminate mass of Internet users whose access and participation may, however, be partially constrained. In addition, weblogs represent specific hypertexts which encourage selective forms of discursive interactions. While some properties of weblogs (e.g. entry arrangement, hyperlinks, HTML/XML code) are stable, others continue to come in great varieties, such as quality, content, purpose and ambition. (Walker 2005: no pagination)

## 2.2 The Formal Composition of Weblogs

With the help of the preceding working definition composed in the last section, some of the major formal parameters and technological features of weblogs can now be illustrated. Generally speaking, a weblog's text design can be expressed on two interrelated levels. The first compositional level provides general information on the weblog, e.g. its author(s), content(s), purpose(s), thereby framing and securing user expectations. The second level represents a weblog's centre stage, hosting its consecutive entries and comments. While we may call the first level, the *contextual plane* as it sets the ground for a weblog's communicative exchange, we could name the second level *discursive plane* as it maps the discursive interaction generated upon the weblog:

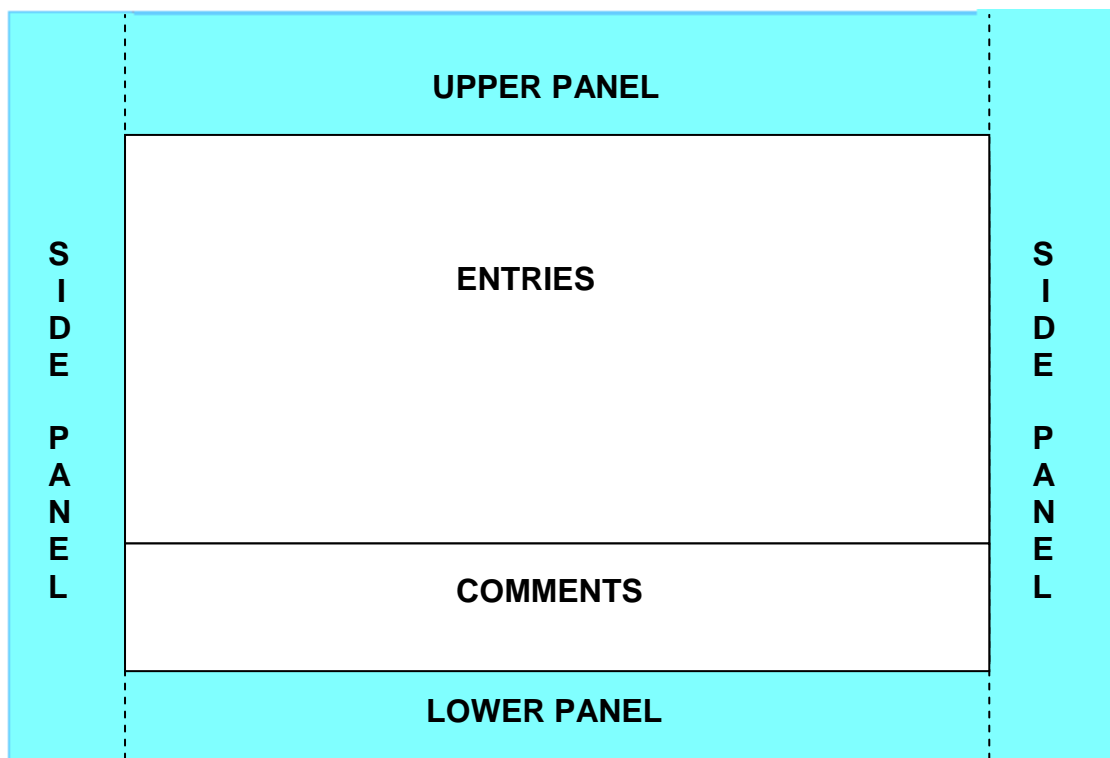


Figure 1. The Main Compositional Spheres of a Typical Weblog Template

The first compositional level of a weblog's text design comprises so called *panels*, prefigured templates named according to their relative position on the computer screen. We can distinguish upper, lower and side panels based on their related spatial position on the computer screen. While these panels collaboratively construe the visual backbone of weblogs, each single one has its own medial affordances in relation to space, time and semiotic deployment and also answers to its own internal compositional arrangement. The latter shall now be inspected, revealing the panels' individual formal organization as well as their typical semiotic and technological structuring.

### 2.2.1 The Upper Panel

The upper panel represents the weblog's masthead. Similar to the design of newspapers, it usually consists of a main title and/or subtitle as well as some form of iconic rendition which underscores the thematic orientation of the weblog or establishes visual clues which point toward the blogger's virtual identity. Regularly, the titles are displayed in fonts of various types or sizes which correspond to their visual and textual prominence on screen. Typography thus often acts as an additional identity trait and reinforces a reference made to a corresponding picture or iconic image set in the background of the panel. See below for a number of examples of topical framing and a concurrent establishment of the blogger's web identity via typographical means in weblogs' upper panel:

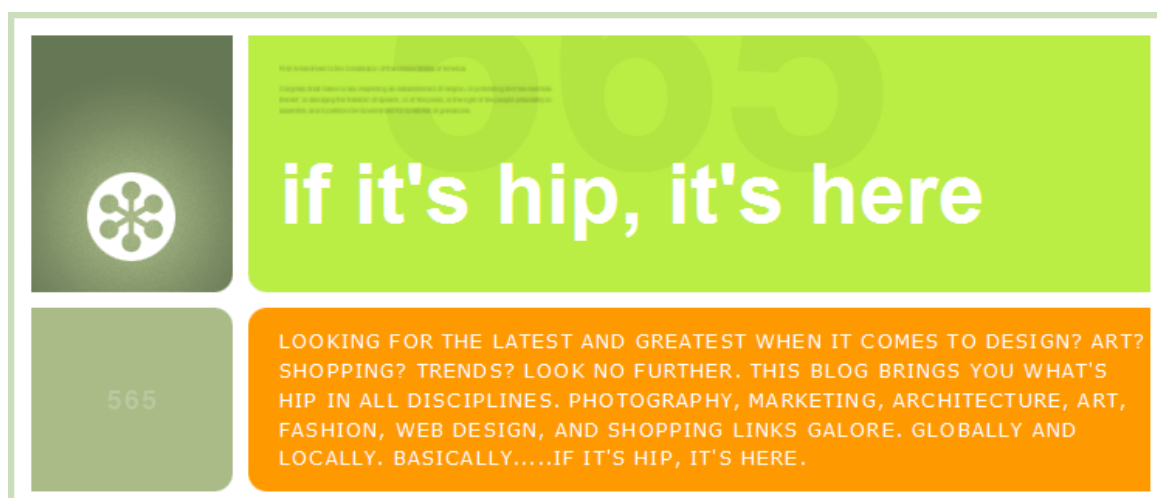


Figure 2: Upper Panel with header and subtitle framed by abstract, conceptual images  
(<http://ifitshipitshere.Weblogspot.com/>)



Figure 3: Upper Panel: A blend of photographic and pictorial icons to stylize header and subtitles  
[\(http://kariandkijsa.Weblogspot.com/\)](http://kariandkijsa.Weblogspot.com/)



Figure 4: Upper Panel: Horizontal Alignment of Verbal and Photographic Information  
[\(http://oneminutewriter.Weblogspot.com/\)](http://oneminutewriter.Weblogspot.com/)

Most weblog services allow bloggers to insert a main title (*header*) and a subtitle into preordained frames centred above the entry section. Both, main title and subtitle are usually depicted consecutively, with the main title assuming visual salience via large size fonts, typographical means (e.g. serifs, italics and bold scripts) and a number of spacing and framing techniques. The main title is typically positioned in prominent centre position within the panel. All of these design resources collaboratively enhance the visual salience of the main title as a primary identification and orientation device for weblog users. As opposed to the main title, subtitle information is normally relegated to marginal areas of the frame, depicted in smaller size, also using paragraphs, lines, colours and boxes to dislocate the subsection from the main title (see figure 6). The additional information contained in subtitles usually fulfils at least one of two primary functions: it either elaborates the main title itself or it reveals the prevalent topic or purpose of the weblog. Pictorial elements are either placed in the background, elevating the verbal information depicted in different colours to set it

apart from the underlying iconic sphere.<sup>13</sup> Background pictures or photos are used to either set the tone and “style” of the weblog by portraying abstract or nebulous images, focussing on colour, modulation and graphic composition of pictorial elements (see example above). They may, however, also be quite naturalistic, referring to specific locations (in travelblogs), occupations (professional Weblogs), persons or character traits (personal or fan Weblogs), companies (corporate Weblogs), objects (hobby Weblogs). This is done to evoke a certain feeling or subscribe to a specific ideological stance, besides strengthening the verbal identification purpose of the main title.

### **2.2.2 The Side Panels**

Side panels include a number of navigational tools necessary for the organization and access actual weblog content. Such internal and external navigation within as well as beyond the weblog is realized with the help of the following tools:

- search engine
- blog roll (list of hyperlinks to related weblogs)
- weblog archive (in chronologic or thematic order)
- tags / tag clouds (listing pervasive themes or categories)
- trackbacks (specific hyperlinks indicating that a particular entry has been quoted in another weblog)
- recent comment section (a quick table summarizing the latest comments posted on the weblog).

In addition, side panels provide other technological features which serve to attribute specific self-images to bloggers. (Virtual) book shelves (bloggers’ personal choice of books) or “about me”-sections, by which weblog authors introduce themselves and present their blogging objectives), may serve this general function (cf. figure 5).

---

<sup>13</sup> Sometime it is difficult to determine which pictorial elements have been chosen from prefigured templates offered by a weblog service and which ones have been created by bloggers themselves. Still, the decisive act of choosing a visual template from a number of other possible backgrounds already is a deliberate choice within the logic of bloggers’ virtual identity.

		RECENT COMMENTS  Berlin 6 Open Access Conference » Wrapping up Berlin 6 on A few thoughts on the heels of Berlin 6  On enabling OA « Be openly accessible or be obscure on A few thoughts on the heels of Berlin 6  Fortune 500 corporate blogging - Nike, SAP and Texas Instruments on SAP's response to my criticism  PR Communications on What corporate blogs look like: JNJ, Chrysler, Palm, Marriott  PR Communications on Why don't we put the Google N-gram corpus on the Web?
<b>Archives</b> » February 2008 » January 2008 » December 2007 » November 2007 » October 2007 » September 2007 » August 2007 » July 2007 » June 2007 » May 2007	<b>Categories</b> » General » Technical  <b>Links</b> » Semantics & Pragmatics » eLanguage News » Open Access News » semantics etc. » David Beaver » Kai von Fintel	

Figure 5: Different Side-Panels Elements (Navigational Tools and Menus)

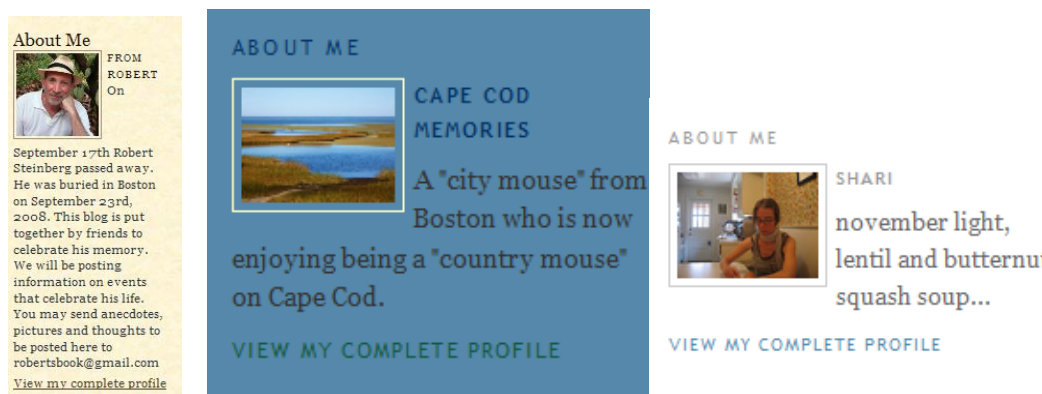


Figure 6: Three Different “about me”-Sections  
(<http://doctor-chef.Weblogspot.com>, <http://capecodmemories.Weblogspot.com/>,  
<http://theglassdoorknob.Weblogspot.com/>)

## 2.2.3 The Lower Panel

Lower Panels of weblogs usually contain copyright information about weblog content, affiliations as well as software-related information and few navigational tools. Sometimes lower panels retain the graphical framing induced by upper panel background motives, thereby emphasizing the compositional distinction of two levels, i.e. the contextual and discourse plane.

Figure 7: Software-Related Information in the Lower Panel of a Weblog (www.jill/txt.net)

Although there is, of course, also “content” in side panels, the main discourse of weblogs takes place on this second level, arranged in the centre part of the weblog. Visually, various weblog entries are placed in reverse chronological order with the most recent entry placed first. Entries are distinguished by various text design resources, such as lines, symbols or icons as well as certain finishing statements. Each entry is accompanied by a hyperlink which, if activated, reveals a series of commentaries. The latter are then often visually attached to the actual entry as illustrated in figure three. Entry and comment length and number may vary with respect to discourse purpose, theme and popularity of the actual weblog.

## 2.2.4 The Entries

Entries are usually initiated by an indication of the date on which the respective text was posted to the weblog. In second position features the title of the entry which can range from single noun phrases to elaborate sentence patterns. Graphical emphasis on the title is typically realized by bold or italic fonts, the use of brackets or a change of font size in relation to successive text parts. Thirdly, entries consist of a main text which might be split up in different paragraphs when exceeding more than five sentences (approx. 100 words).

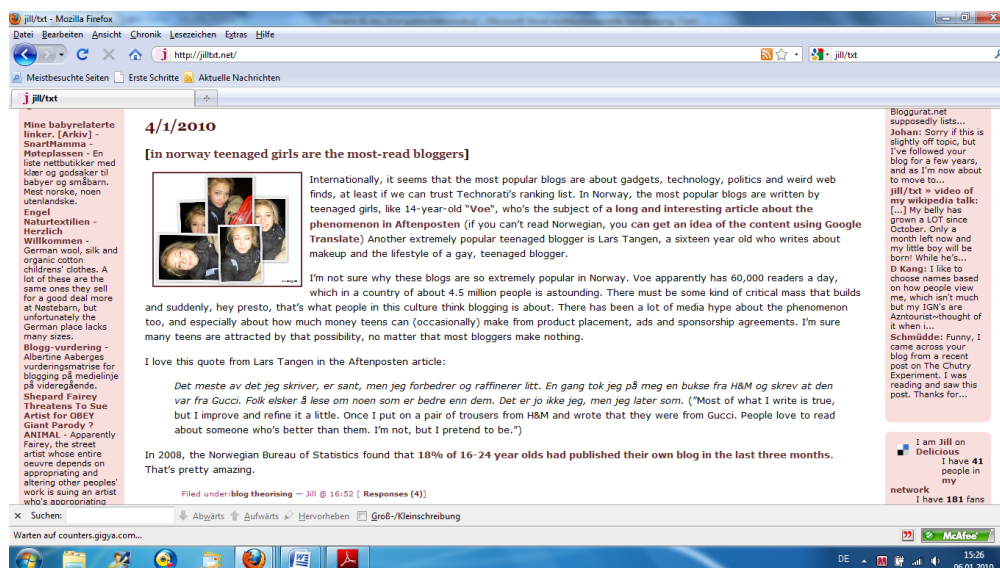


Figure 8: Prototypical Weblog Entry with Main Title, Main Text and Hyperlinks (www.jill/txt.net)



The main text as well as the main title of an entry may include various embedded hyperlinks leading to further internal or external text resources. Also, pictures or photographs may be included into the main text of an entry. According to relative size, the iconic image can be positioned before, in between or after the main text, often carrying a respective caption to identify and related the picture. Sometimes, small-size pictures are embedded in the textual narrative, lacking an individual caption.

### 2.2.5 The Comments

Commentaries are usually shorter than their related entries. Mostly, they directly respond to issues voiced in the entry, evaluate opinions or add further information to the weblog. For the latter purpose, hyperlinks to other Internet resources or weblog posts may be created in comment sections. Iconic images cannot be represented in comments. They are thus often replaced by emoticons or by hyperlinks leading to iconic resources on other Internet sites. On a semiotic plane, however, comments entail strictly verbal content only. Weblog authors<sup>14</sup> may engage in comment sections and stimulate multiple conversational exchanges. As opposed to this author status, weblog users may respond to each other in comment sections but are prohibited to write their own entries. Each comment exhibits a particular time-stamp indicating the moment of its transmission to the weblog.

#### 2 Responses to “digital knitting”

1. *Kjerstin* Says:

November 3rd, 2008 at 12:21

Oh, this is great! I rediscovered knitting last year and have been looking for a simple way to make pictures into patterns (more simple than using grid paper, that is) ever since. There's just no better November activity :-)

2. *Ellen* Says:

November 4th, 2008 at 16:58

I've been trying to work out the appeal for me. Part of it is that it functions so well as a database--this really appeals to the researcher in me (after all, I'm an academic, for better or worse). The sheer volume of information (patterns, yarns, projects, people) is impressive. Then part of it is the voyeur/exhibitionism aspect. I love getting glimpses into people's generally domestic lives--family, home, private pass times. And of course it is frighteningly fun to expose my own private life. I find myself totally okay with posting images of myself and even my family members there, which I probably wouldn't do anywhere else. Finally, there is something really satisfying about documenting my knitting--the stash, the works in progress, and the finished objects. My husband compares it to my obsession with my CV--he wonders if sometimes I write things just to be able to add another publication to my list, and it feels the same way with Ravelry. And it's not just the amount of projects, but the overall aesthetic effect of my project page that I think now is starting to drive which projects I choose to make. Oh, and of course being able to "friend" people, join in discussions and trade and purchase things from real people in the community makes it a very rich social experience as well. I definitely think Ravelry has made a difference in what and how I knit!

Figure 9: Comment Section Including Two Consecutive Weblog Commentaries (www.jill/txt.net)

<sup>14</sup> At this point, it seems necessary to define on which basis weblog authors will be delineated from users. I hold that weblog authors are registered authors with access to the composition and deletion of weblog entries and comments. Weblog users may contribute content to weblog comments and thus turn into viable “authors” of some kind. They, however, will not reach the status of veritable weblog authors who exhibit the following qualifications: 1. General Authority, 2. General Accessibility, 3. General Addressability, 4. Responsiveness, 5. Regular Activity.



Following this first general classification of weblogs' main compositional parts, segments and features, we shall now return to the question of their technological, formal and functional genesis and function in discourse.

## The Genre

### 3.1 The Naturalization of the Blogosphere

Weblogs are multi-dimensional phenomena whose conception touches upon four adjoining conceptual scopes. These spheres reflect weblogs in their material basis (*medium*), in their composition (*application*), in their purpose and topic (*genre*<sup>15</sup>) and in their linguistic form (*text type*). Realize that one scope does not attain any kind of prominent status. Each one is heavily contingent upon the others, so that the emergence of new forms of communication is the result of an ongoing interaction between the various scopes. In taking with this gradual approach, we may also retrace the various conceptual paradigms that moulded communication on the Internet throughout the last decade

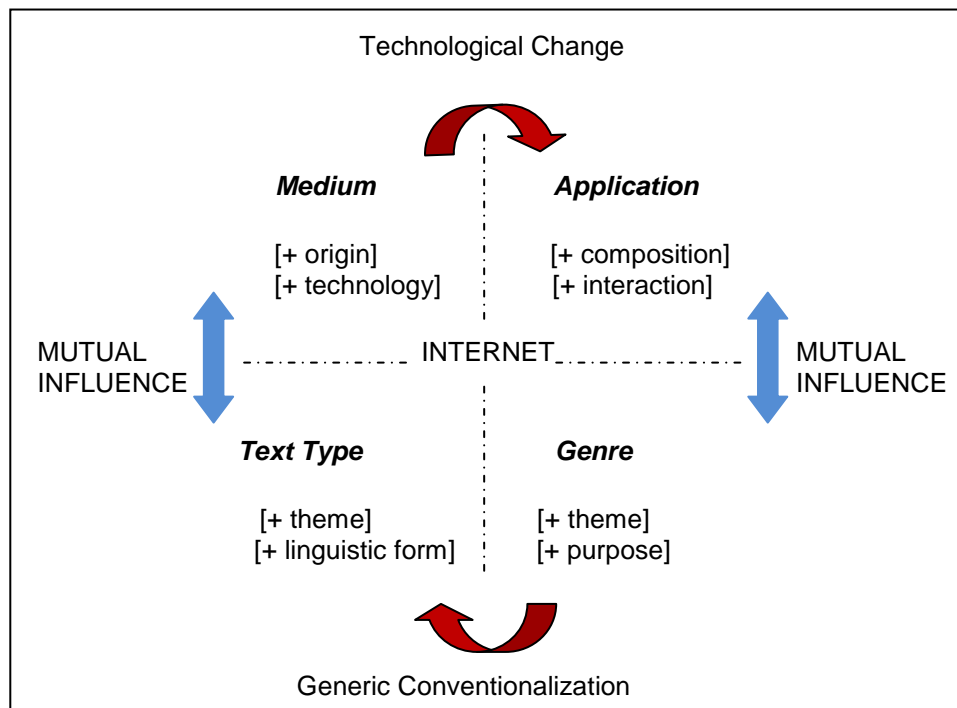


Figure 10. The Genrefication Cycle of Computer-Mediated Communication

On the basis of these conceptual scopes in figure ten, we may stipulate four chronological stages in the development of weblogs:

<sup>15</sup> For definitions and an elaborative discussion of the concepts of *text genre* and *text type*, see chapter 3.2.

### **Stage 1: *Simulation* (1996 - 1999)**

In a first stage, weblogs did not strictly emancipate from web pages. As such, Weblogs were no more than meta-sites, displaying regularly updates of annotated link lists aimed at facilitating information retrieval for a restricted group of tech-savvy individuals. This first phase of development only started to experiment with independent blogging form, greatly relying on website technology and composition.

### **Stage 2: *Specialization* (1999 - 2002)**

A second, cumulative stage set in 1999 when bloggers increasingly appropriated old media genre conventions from a limited set of old media text genres, largely epitomized by the vast usage patterns borrowed from written diaries. This second phase spawned a crucial process of formal generic independence, establishing Weblogs as original web applications with specific compositional traits and themes different from their technological forerunner, the website. The process equally gave birth to a renewed “personalization” of blogging themes, centring on the individual life and perspective of the blogger rather than on more general and objective vantage points.<sup>16</sup>

### **Stage 3: *Stratification* (2002 - 2005)**

With users increasingly accessing and generating weblogs, formal and functional variation of weblogs soon blossomed. This development soon changed the image weblogs as electronic spin-offs of the classic diary. Different weblog types started to branch off from the traditional line. At the same time, McNeil (2009) reports that bloggers’ begin to discuss the motives of their blogging practice. To this end, debates on what purpose weblogs should serve mushroomed on the Internet. In these discussions, bloggers usually drew on their individual associations of classic text genres to describe what weblogs should be about. Thereby, they associated or dissociated different bloggers and their discursive works in the blogosphere. Two pervasive blogging “camps” emerged from these discussions. On the one hand, camp one rejects the concept of weblogs as personal diaries. On the other hand, camp two rejects the vision of weblogs as a form of “citizen journalism” (cf. Miller &

---

<sup>16</sup> This second stage was, of course, induced by a technological change, i.e. the growing availability of weblog services with which regular users could easily publish their own weblog without necessarily requiring an in-depth knowledge of programming languages, such as HTML or XML. This opened up the possibility of creating a free personal writing space online for everyone with access to a network computer.

Shepherd 2009).<sup>17</sup> The general image of the Internet diary is resonating in clichés of unprofessional private diaries which are supposedly composed by (mainly female) writers. The latter are expected to disclose their personal experience to only few privileged insiders. Such extreme views of what weblogs should not represent are further consolidated by representatives adhering to the second camp of bloggers. It insists on the image of weblogs as professional public journals (as opposed to diaries) created by predominantly male blogging elite (cf. Herring & Paolillo 2006). Weblog authors reclaim “actual blogging status” on the basis of an allegedly more “objective”, quasi-professional stance, reporting on matters of general interest rather than personal experience. Obviously, bloggers, in this radical sense, posit a stark contrast between their own blogging agenda and the one of Internet diaries.<sup>18</sup> Professional weblogs, it is believed, needs to reflect public rather than private issues. This topical contrast clearly delineates them from the “old and passé” of the classic diary format. In actual practice, however, McNeill (2009) concurs that such rigid distinctions do not hold. In fact, possible candidates for the private diary type have always moved beyond their private domain just as professional weblogs continue to profit from a paramount sense of personal authenticity which is clearly derived from user expectations originating in the diary genre.

#### **Stage 4: *Interaction* (2005 - 2010)**

In the last years, bloggers and researchers alike have begun to focus their attention on the participatory potential attributed to the recent web tools often referred to as *Web 2.0* technology (O'Reilly 2005). The *ripple effects* of interactive network tools, e.g. *comments*, *permalinks*, *RSS-feeds*, *chat boxes* are perceivable by the day with bloggers and weblog users making increasingly use of technologically-enhanced strategies for textual collaboration. Blogging definitions thus typically adhere to weblogs' current dynamic shifts between “semi-active” reading interaction and “fully-active” participatory text collaboration. What O'Reilly (2005) calls “harnessing

---

<sup>17</sup> Blood (2002) advocates a similar set of weblog genres. She portrays three different weblog types, i.e. filter, notebook and personal Weblogs, which are equally divided by recourse to their relative position on a conceptual private-public divide. Filter weblogs are, however, somewhat different to the concept of McNeil's “weblogs” which affirm the salient compositional and technological properties of weblogs; filter blogs refer to weblogs first evolutionary stage (see Barger 1997) and do not comprise specific compositional traits.

<sup>18</sup> The struggle of defining the weblog is especially pronounced in the blogosphere itself where “metablogging” (Trammel & Keshelashvili 2005) or “meta-genre talk” (Giltrow 2002) is essentially constrains the formation of weblog composition, structure and usage over time. Hitchcock (2005:203) reminds us, in this vein, that bloggers deliberately act as “gendarmes of genre” to police what is right and what is wrong in their texts.

collective intelligence” is another productive generative strand of the current blogosphere. With more and more bloggers collecting, sharing, tagging and distributing information collectively singular weblogs are gradually growing into collective networks of shared interests and information. There is reason to believe that the technologically-enhanced weblogs of tomorrow will be extended into so-called social-network sites, the veritable offspring of the blogging movement. However, it is the current preliminary status of weblogs as restrictive zones of personal engagement and public portals for friendship building which characterizes today’s weblogs. Therefore, some definitions explicitly emphasise the private-public dimension as being particularly vital to weblogs formal as well as functional role in CMC. As shown elsewhere for websites (cf. Eisenlauer & Hoffmann 2008), these four elemental stages of weblogs’ generic naturalization may also be expressed in the terms of remediation theory (Bolter & Grusin 2003). The latter refers to the process of “generic exchange” between old and new media, one borrowing from and enhancing the other:

In Western Europe the shift from handwritten codex to printed book was another such refashioning, and the shift to electronic writing is yet another. We might call each such shift a “remediation” in the sense that a newer medium takes the place of an older one, borrowing and reorganizing the characteristics of writing in the older medium and reforming its cultural space.

(Bolter 2001:23)

According to Eisenlauer & Hoffmann (2008:4), the generic evolution of CMC forms of communication acquire conventional forms according to intermedial exchange patterns, i.e. *simulation*, *improvement*, *refashioning* and *incorporation*. Using figure four as a base model, we may align the aforementioned chronological stages of the blogging movement with the four primary phases of remediation. Whereas in the *simulation* phase weblogs fully rely on the appropriation of formal and functional conventions of other text genres, in *improvement* weblogs show first signs of generic independence, such as the surfacing of comment sections or first compositional traits (Stage 1). Greater generic autonomy arises in *refashioning* when a certain level of compositional and thematic uniformity is reached and differences rather than similarities between weblogs and similar forms of communication, e.g. websites get emphasised (Stage 2). The *incorporation* phase sees weblogs implementing a range of different forms of expression into their text design resources, for instance weblogs using film, pictures, photos or other semiotic resources to deliver a certain message.

At the same time, text conventions from a various old media text genres are embedded in the generic interface of the weblog (Stage 3 & 4). From the backdrop of these evolutionary patterns and their relation to technological, formal and functional characteristics of weblogs, we may detect important criteria central to various weblog genres. These shall now be elicited in a next step.

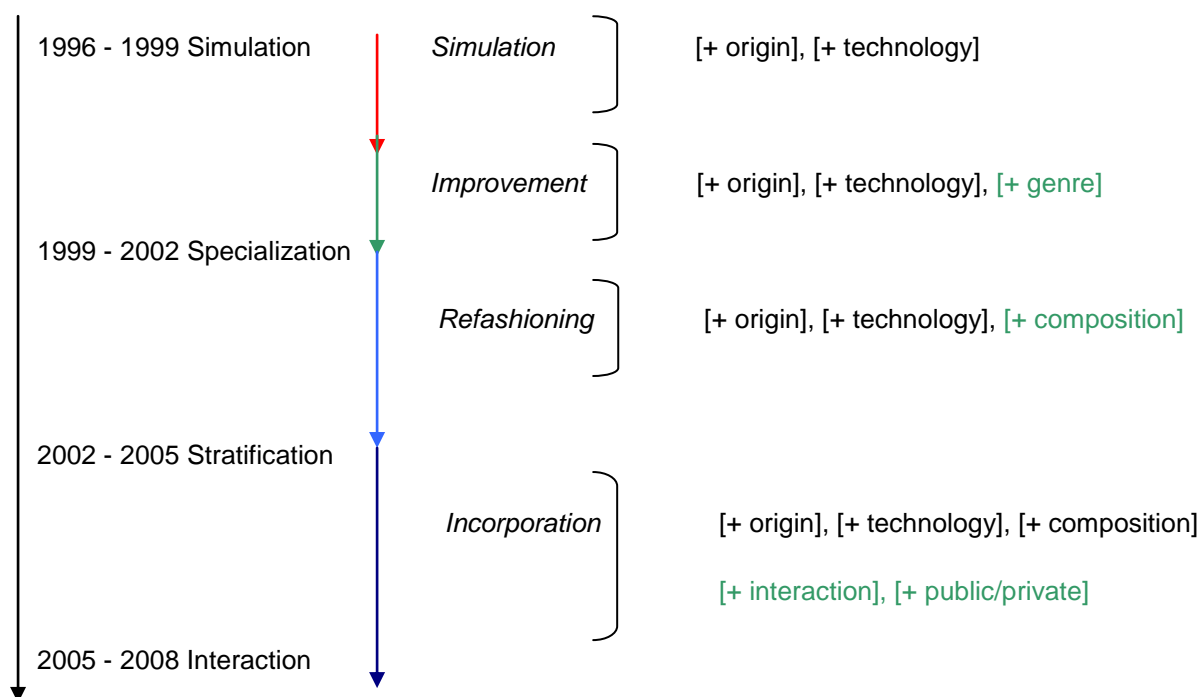


Figure 11: Evolutionary Stages and Descriptive Properties of the Blogosphere

### 3.2 Diary, Journal or Weblog? The Path Toward Generic Attribution

Generic Research on Internet-based forms of communication has become an important field in linguistic CMC research. In recent times, genre theory seems to have risen like a sleeping giant from the panoply of stirring issues surrounding weblog-related research. Yet, while Internet users readily proclaim the advent of “new” Internet genres, pointing to their specific technological and formal appeal, web-based discourse is prone to proliferate into various, multi-functional text types, thereby resisting precise generic attributions. As shown above, new media genres borrow typological patterns from “old media” only to repurpose or, more aptly, to *remediate* them in a process of consistent naturalization (Eisenlauer & Hoffmann 2008:6). It follows that an early appropriation of the diary format was quickly responded by a systematic enhancement of formal and functional weblog features moving toward further generic differentiation (cf. Ong 2003/1988, Crowston &

Williams 2000, Dillon & Gushrowski 2000). The latter process heightens the need for generic clarification.

To understand this apparent “contrast between the perceived discourse reality of Internet users and the results of CMC research” (Heyd 2009:241) we must take into account what *genre* actually means in this context. In linguistics, genre studies recurrently draw on the concept of *genre* as detailed in Swales (1990). He finds text genres to

[...] comprise [...] a class of communicative events, the members of which share some set of communicative purposes. These *purposes* are recognized by the expert members of the parent discourse community and thereby constitute the rationale for the genre.  
(Swales 1990:58, my emphasis)

According to Swales’ definition, text genres refer to the communicative purpose or goals of textual entities. The emphasis on discourse purpose as the central criterion for generic classifications also reverberates in other linguistic definitions of genre (cf. Painter & Martin 1986, Bathia 1993). What remains unresolved in these descriptions is the relationship between discourse purpose and structure. Some linguists insist on the fact that communicative purpose and structure need to be kept apart for epistemological reasons. Others proclaim that both are necessarily coupled in meaningful discourse (cf. Ventola 1987, Lemke 2005).<sup>19</sup> Others dispute the existence of a symbiotic interdependence between textual form and function, showing that discourse function may trigger various textual forms, while linguistic form may not necessarily amount to one, clearly dissectible underlying text function:

There is no neat fit between sociological and linguistic categories [...]. One cannot, it seems, have it both ways with language. Either theory or method are formally neat but semantically messy or they are semantically neat but formally messy.  
(van Leeuwen 2008:24)

In order to illustrate the discrepancy indicated in the above quote, let us consider, for illustrative reasons, the genre of *fiction*. It comprises numerous sub-genres, e.g. fairy tales, short stories or various sorts of novels all of which differ in regard to their individual linguistic form and composition. It should therefore be rather difficult to spot

---

<sup>19</sup> According to this systemic-functional view of genre, purpose and form combine into motivated, recurrent patterns, or as Martin & Williams (2004:122) have it, genre “is a high level configuration of meanings responsible for describing the ways in which a culture goes about combining field, mode and tenor and phasing them together in discourse.”

the co-presence of linguistic and/or compositional traits in all of these text forms. All we can hope to find are family resemblances between the different kinds of fictitious discourses. As a result, most linguists are cautious enough not to expose specific compositional and/or formal criteria as stable qualities of text genres. For Swales (1990), descriptive fusions of functional and structural criteria merely amount to what he puts as “high probability expectation[s]” (Swales 1990:58). In much the same vein, Biber (1988:70) claims that “in a fully developed typology of texts, genres and text types must be distinguished, and the relations among them identified and explained”.<sup>20</sup> He goes on to define *genre* “on the basis of external criteria”, notably relating to the speaker’s purpose and topic”, keeping to Swales classic outline of the notion. On the other hand, he uses the notion of *text type* to refer to “groupings of texts that are similar with respect to their linguistic form, irrespective of genre categories” (1988:70). Other linguists typically agree with this approach, like Schubert (2008) who reaffirms Biber’s segmentation of the two textual categories:

Texttypen (*text types*) sind Einteilungen von Texten in eine stark begrenzte Anzahl von Kategorien auf hoher Abstraktionsstufe und auf Basis ausgewählter linguistischer Kriterien. Genres (*genres*) dagegen beruhen auf alltagssprachlichen Bezeichnungen von Textsorten, weswegen sie sehr zahlreich sind und eine niedrige Abstraktionsstufe aufweisen.  
(Schubert 2008:89)

[*Text types* organize texts into a very limited number of categories on a high plane of abstraction and based on selective linguistic criteria. In contrast, *genres* rest on common, everyday terms for texts which is why they are so numerous and exhibit a low degree of abstraction.]  
(Schubert 2008:89; my translation)

In line with Schubert’s stance, I shall adhere to the useful distinction of the concepts of *text genre* and *text type* throughout this study. *Genre* refers to a categorization of texts on the basis of thematic and functional affinity, and applying the notion of *text type* to text classifications on the basis of recurrent compositional and/or linguistic patterns. Applying these two definitions to Weblogs, we realize that a general, super-genre of Weblogs encompassing their multi-functional breadth and thematic diversity seems hardly attainable. Indeed, I already explicated the related conceptual drift of Weblogs from generic identification to dispersion, from monologue to polylogue and

---

<sup>20</sup> Systemic-functional linguistics also adheres to the distinction of two quite similar categories, i.e. genre and register. The term *register* is used to address classes of text which share similar linguistic (lexico-grammatical) patterns. For more exhaustive information on the various approaches and terms related to genre theory, see Bathia 1993.



from private to public discourse in the previous chapter. Therefore, communicative purposes, which principally determine generic orientation, seem “highly stable elements of discourse [and are] relatively robust against medial change” (Heyd 2009:241). Searching for radically new text genres in CMC therefore does not appear to be very promising. The transient generic nature of Weblogs is, however, not in any way exclusive to Weblogs. It is likely that all Internet-based forms of communication undergo a similar path towards their ultimate goal of generic autonomy:

It is likely that all Internet genres – classic example: the weblog [...] are undergoing a process of eking out an identity between a pedigree of genres in the traditional media and the technical constraints of the new medium, which is itself in a process of permanent technical evolution.  
(Stein 2006:3)

Briefly, we may not be able to define the weblog as a CMC genre per se but as software applied to the end of composing various different weblog genres. Hence, we may after all be capable of describing what Heyd (2009) calls functional ‘supergenres’ as well as various ‘subgenres’ which, in turn, may exhibit minor formal and structural differences. To identify these ‘supergenres’ on the grounds of blogging functions, I shall draw on Technorati’s state of the blogosphere report conducted in 2008. Its results reveal bloggers’ respective blogging purpose(s):

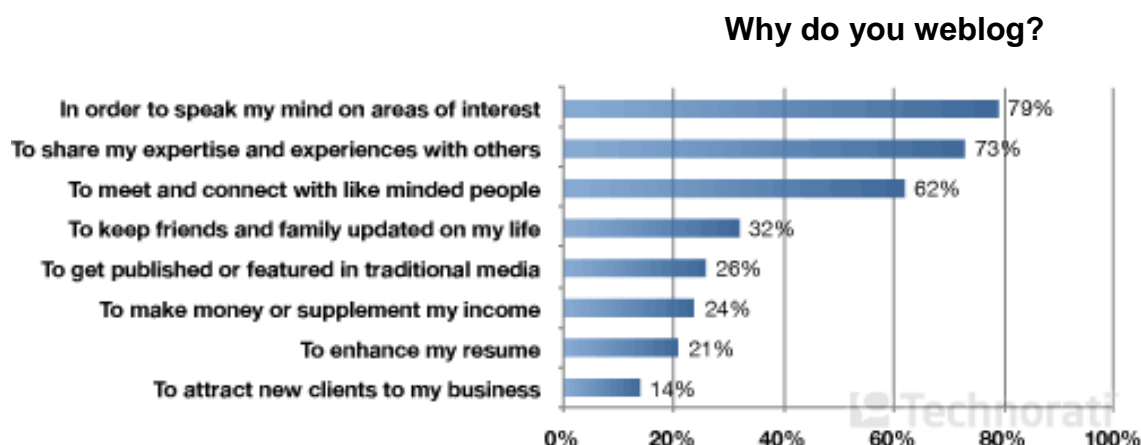


Figure 12: Common Blogging Purposes of Bloggers (www.technorati.com)

The weblog functions elicited in previous figure can be condensed into the following salient blogging purposes:

- Self-Expression
- Self-Promotion
- Friendship Building

- Networking and Organization (Work-Related Blogging)
- Commercial Interests.

Again, on the basis of these primary blogging interests, we may now delineate four preliminary genres and a set of interrelated subgenres.<sup>21</sup> The genres are

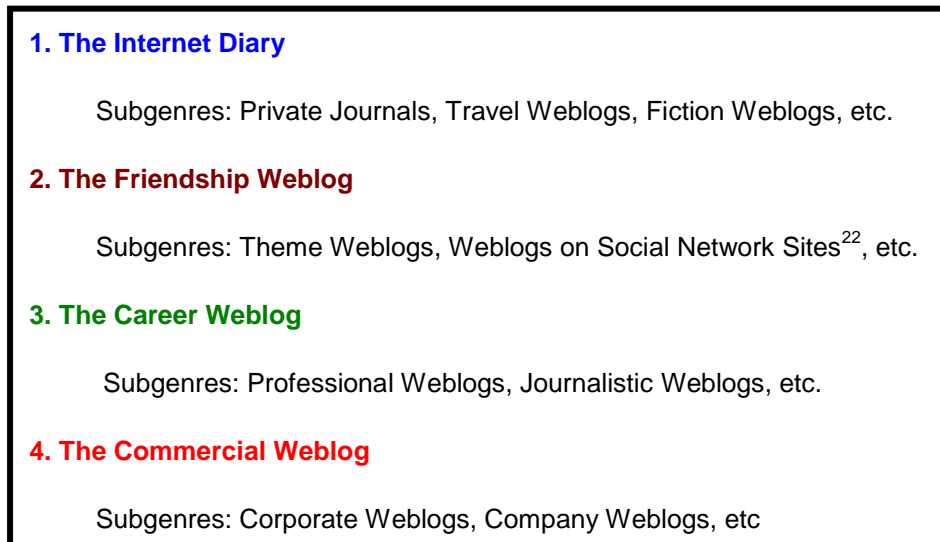


Figure 13: Four Salient Weblog Genres on the Internet

Similar to this set of genres, previous research already proposed similar weblog types as early as Blood (2003:17). She distinguishes between three general weblog types called *filter weblog*, *notebook* and *personal journal*. Herring et al (1994) summarize their main characteristics as follows:

The content of filters is external to the blogger (world events, online happenings, etc.), while the content of personal journals is internal (the blogger's thoughts and internal workings); notebooks may contain either external or internal content, and are distinguished by longer, focused essays. (Herring et al 1994:2)

Whereas Blood's weblog type descriptions seem either dichotomous, rigid (filter weblog vs. personal journal) or fuzzy (notebook), the functional distinction of my own weblog genres follows a prototypical identification of discourse purpose (cf. Rosch 1978, Wittgenstein 2004/1953). According to Wittgenstein (2004/1953) entities which are said to belong to the same class do not necessarily share the same set of

<sup>21</sup> Note that some subgenres mentioned may possibly be listed in more than one genre; a fact which pays tribute the generic flexibility of weblogs and reflects the epistemological, rather than "real" status of this classification.


<sup>22</sup> SNS Weblogs refer to weblogs as implementations to Social Network Sites, e.g. facebook®, StudiVZ®, Orkut®, etc.

properties. Rather some entities may share one set of characteristics, while other ones may share a different one. All are therefore connected by a certain sharing family resemblance between each other:

Why do we call something a "number"? Well, perhaps because it has a direct relationship with several things that have hitherto been called number; and this can be said to give it an indirect relationship to other things we call the same name. And we extend our concept of number as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres.

Wittgenstein (2004/1953:39)

In this vein, weblogs do not all share the same properties but describe a category which is “interconnected by similarities between its members” (Löbner 2007:178). The same actually holds true for the aforementioned weblog supergenres. Furthermore, they embrace blogging functions to various degrees, exhibiting more central and more peripheral weblog purposes. Figure three exposes these dimensional attributions by degrees of membership. On its basis, we may allocate specific weblog examples to supergenres, allowing for *fuzzy boundaries*:



<b>Proto-typicality</b>	<b>Internet Diary</b>	<b>Friendship Weblog</b>	<b>Career Weblog</b>	<b>Commercial Weblog</b>
1.	Self-Expression	Friendship Building	Networking	Commercial Int.
2.	Self-Promotion	Self-Promotion	Commercial Interest	Self-Promotion
3.	Friendship Building	Self-Expression	Friendship Building	Networking
4.	Networking	Networking	Self-Promotion	Friendship Building
5.	Commercial Interest	Commercial Interest	Self-Expression	Self-Expression

Table 3: The Prototypical Functions of Weblog Genres (1= Central, 5= Peripheral)

This model of a dimensional genre attribution embraces Erickson’s concept of *genre ecology* (Erickson 2000).<sup>23</sup> It helps to locate individual weblogs on a spectrum of possible purposes. The level of prototypical salience can be detected by searching for motivated patterns of linguistic and technological realization. This account clearly contrasts with the dichotomous, ideologically-laden viewpoints of an Internet diary a (filter) weblog which are both laden with ideological deprecations of weblog genres. Whereas the former aligns with wide-spread gender connotations of diaries being

<sup>23</sup> Erickson’s notion of *genre ecology* (2000) refers to “the possibility of genre change and overlap” (Heyd 2009) which accounts for generic family resemblances between old and new media text genres, e.g. the *diary*, the *journal*, the *article* or the *weblog*.

'private', 'restricted', 'spontaneous', 'feminine', 'unprofessional' or 'nonsensical' text forms, (filter) weblogs are allegedly 'public', 'participatory', 'technological', 'reflective' and 'professional'. Such personal misapprehensions emerge as current features of so-called "meta-genre talk" in which most weblog authors are vividly engaged (Giltrow 2002). Although such talk reifies some bloggers' overt attempts to establish generic stability on the Internet, it does not, however, reflect the current, empirical status of web-based weblogs.

### **3.3 Personal Weblogs as a Meta-Genre**

In the course of this study, we shall focus primarily on the two pervasive weblog genres of *Internet diary* and *friendship weblog*. Both can be subsumed to the superordinate genre of the *personal weblog* which functionally adheres to high ratings of self-expression, self-promotion and friendship building, showing comparatively low ratings on networking and commercial interests. Likewise, the personal weblog category thematically endorses personal affairs and experiences while avoiding professional, economical or political discussions. The weblog corpus which was compiled for this study includes ten personal weblogs whose purpose and theme(s) had to be carefully assessed to maintain a sufficient degree of generic stability of the corpus data. To achieve this effect, I used two cumulative methods. One of these two steps comprised the distribution of a digital questionnaire to potential bloggers. This questionnaire consisted of ten central questions, aimed to elicit information from each weblog author about the purposes of their blogging. In particular, the bloggers were asked to fill in a digital questionnaire asking them to rank possible purposes for their engagement in this form of communication. I based the form and style of the questions and response options on compatible queries designed in previous weblog studies such as the ones conducted by *Technorati.com* or the *PEW Internet and American Life Project*.<sup>24</sup> I will now discuss two of the ten questions in more detail to illustrate how bloggers' responses to the questionnaire were evaluated and used to create generic profiles of the weblogs. The remaining questions can be reviewed in the appendix to this study.

---

<sup>24</sup> Both technorati and PEW have repeatedly conducted large-scale surveys on the Internet or via telephone both in the United States of America as well as in other countries. The results of the surveys are publically available on the Internet at the following two sites:  
[http://Internet.pewinternet.org/\(11/15/09\)](http://Internet.pewinternet.org/(11/15/09))  
[Internet.technorati.com](http://Internet.technorati.com) (11/15/09)

One of the questions proved particularly valuable to the purpose-driven classification of potential weblogs. The question asked: “Why do you blog?” and extended a set of potential responses for the bloggers to choose from.<sup>25</sup> These were:

1. to express yourself
2. to entertain people
3. to meet people
4. to document personal experience
5. to store data
6. to make money
7. to share skills
8. to influence people
9. to motivate others.

The bloggers were not only asked to tick the response options they endorsed but were likewise encouraged to rank their answers on a scale from one (‘no, I reject this answer’) to ten (‘yes, I fully concur with this response’). They could also insert replies or explanations of their own.<sup>26</sup> As a result, the answers given by the weblog authors proved to be more selective than otherwise. The following blogging aims all showed a minimum rating average of at least six points:

1. express myself creatively (8,5)
2. entertain people (7)
3. network and meet new people (7)
4. document my personal experiences and share them with others (6)
5. stay in touch with friends and family (6).

All of the above blogging objectives seem to align with our generic scheme of the personal weblog. While purposes (1), (2) and (4) clearly align with the discourse aims of self-expression and self-promotion, (3) and (5) connect to the overall discourse purpose of friendship building.

Compare the low average rankings of the following objectives connected to the purposes of job-related networking and commercial interests:

---

<sup>25</sup> In the original questionnaire the order of responses was obviously changed.

<sup>26</sup> It should be noted, however, that only two of fifteen bloggers chose to make use of this option.

1. motivate other people to action (3)
2. share skills and knowledge with others (3.5)
3. influence the way other people think (3)
4. make money (3,5).

Table four represents the average response ratings of twelve out of fifteen bloggers which showed a considerable degree of consistency in their responses.

The first four responses can be easily connected to the genre of personal weblogs (*to express yourself, to entertain people, to meet people, to document personal experience*). In contrast, the other five responses (*to store data, to make money, to share skills, to influence people, to motivate others*) can be aligned with other weblog genres, for instance with those that centre more prominently on economical issues. The responses to this question *inter alia* revealed that twelve out of fifteen bloggers who participated in the survey endorsed response options one to four, while they predominantly rejected responses five to nine.

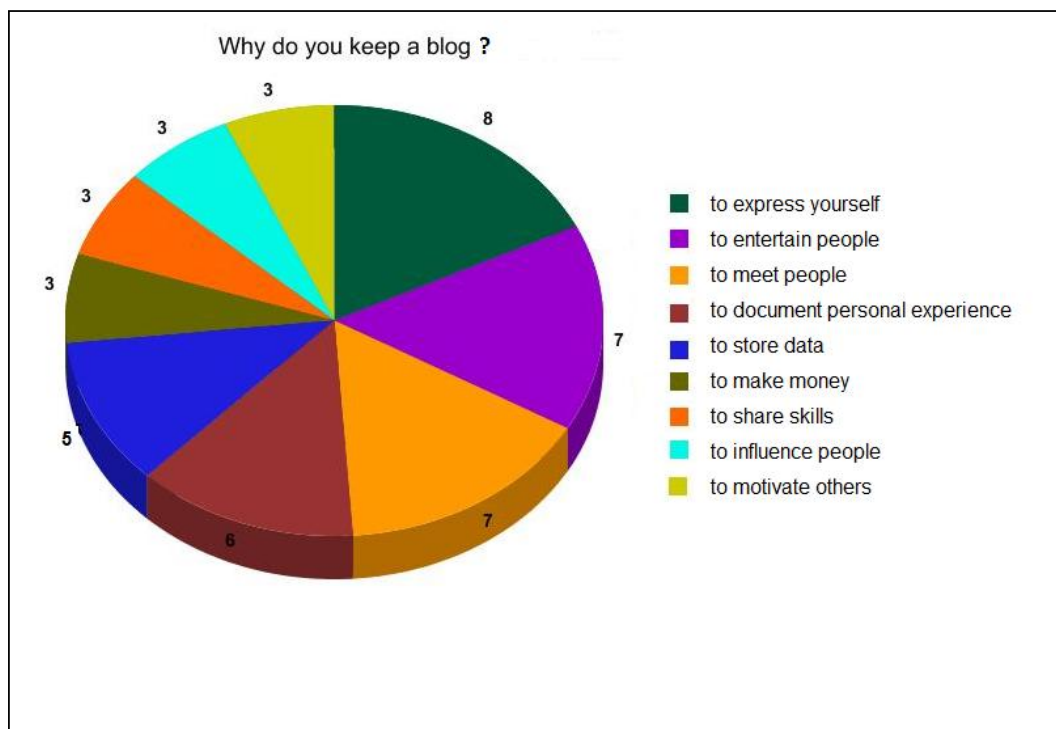


Table 4: Eliciting Blogging Motivations (Ranking Average of Twelve Bloggers)<sup>27</sup>

<sup>27</sup> The numbers dispersed around the pie chart represent the participants' average ratings of the individual blogging purposes.

A second question was posed to secure the generic consistency of the potential weblog candidates. To this end, the bloggers were asked to decide whether they regard their weblog as a private diary, public journal, a friendship base, a link resource or a communication hub. The following figure reveals the mean score of the blogger responses to this query:

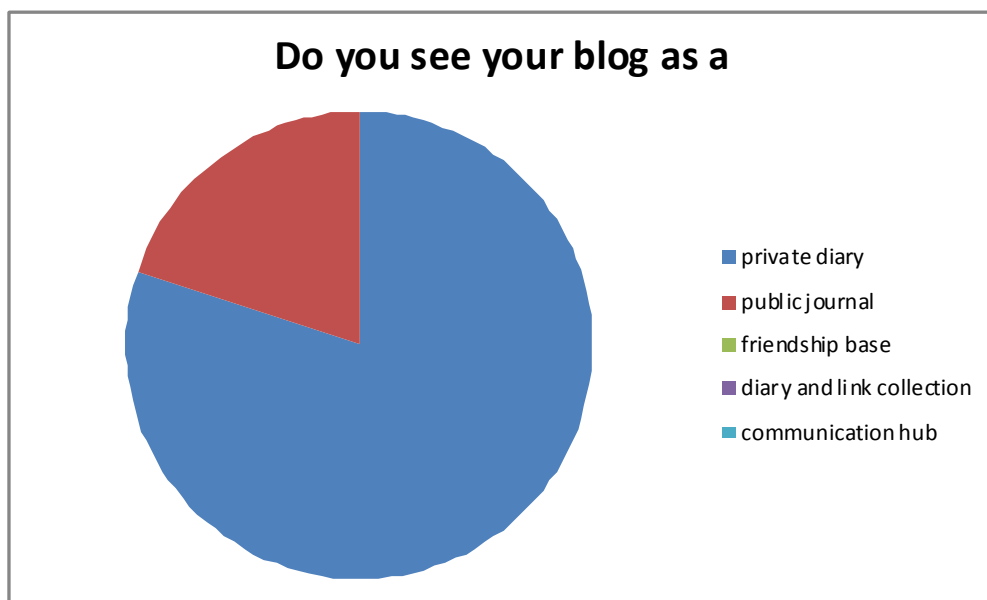


Table 5: Blogger Reflections: Private Diary, Public Journal or Filter Blog?

Again, the preference for private topics (journal) in the public sphere of the Internet is evident. Only two of five possible options were selected by the bloggers, i.e. private diary and public journal. Therefore, bloggers obviously do not wish to refine their blogs as filter-blogs, much less as career or commercial blogs. Therefore, we have reason to believe that the weblogs chosen for this study's corpus share similar discourse purposes which revolve around the expression of personal experience either in the private or public sphere. Their principal aim is to entertain and inform friends and bloggers. The results of the Internet survey thus enable us to maintain a sufficient level of genre stability between the weblogs chosen for this study's corpus. In a second step, a set of tag cloud analyses were executed to assess the topical stability of the weblogs.

Note that participation in the survey was voluntary. Still, the majority of authors (12 out of 15 original authors) responded to the questionnaire. The deliberate

participation of bloggers in the survey turned out to be a crucial step towards determining the salient discourse functions of the weblogs.

#### 4. Mushy's Moochings

Figure 14: Tag Cloud of a Selection of Entries in the Weblog Mushy's Moochings

<sup>29</sup> The visualization of word frequencies was conducted with *Manyeyes*, software which visualizes statistical information in a visual cloud called a *wordle*. The tool can be retrieved at the following URL: <http://maneyeyes.alphaworks.ibm.com/manyeyes/> (27/12/09)



In the course of this procedure, ten out of the twelve weblogs, which had previously passed the first stage of functional selection, could now be considered viable candidates for the AWC.<sup>30</sup> The cumulative results of the functional and topical tests were an important asset in factoring out all unlikely candidates for the AWC and securing the functional consistency of the text data. The two-step methodology of functional and topical assessment via questionnaires and tag cloud operations thus proved successful in establishing generically-consistent weblog data. The process can equally be maintained with a view to other weblog genres and even in regard to other forms of Internet-based communication. It is a great methodological asset in the description of CMC language data and a fair compromise toward building reliable corpus data.

### 3.4 Weblogs as Hypertext

This chapter will provide a general insight into what a hypertext actually is and reveal the key changes it brings to the creation and interpretation of weblog texts. First, we take a look at how printed hypertext can be distinguished from “new”, electronic hypertexts. Then, we elicit of each kind the key characteristics of *hyperwriting* and *-reading*, i.e. the creative and hermeneutic exigencies of electronic hypertext. On the one hand, we learn how hypertext authors strive to accommodate the interests of potential users by modularizing and connecting their texts in interrelated fragments. On the other hand, we investigate how users attempt to make hypertexts coherent, relying partly on their individual pool of (different types of) knowledge and partly on information provided by the text itself.

In summation, this chapter highlights a range of typological differences between medial, textual and cognitive conceptions about hypertext and discusses their technological and linguistic utility for this study. I have already indicated that weblogs are in many ways similar to traditional web pages. Indeed, Rebecca Blood (2003:8) points out that only a decade ago “weblogs could only be created by people who already knew how to make a website”. Today bloggers can fall back on easy-to-use software to create their own weblogs. The software<sup>31</sup> they use for this purpose has evolved from *hypertext markup language* (HTML), i.e. websites’ classic computer code. Therefore, weblogs are (as illustrated in the previous chapter) indeed the

---

<sup>30</sup> Both of these procedures were already presented in more detail in chapter two.

generic offspring of web pages, and I hold that a basic knowledge of how websites are built, maintained, operated and interpreted is highly relevant for any analysis of weblogs. Toward this goal, one will need to get acquainted with the elementary organizing principle of website technology, hypertext.

### 3.5 Three Perspectives on Hypertext

Much academic work has yet been dedicated to the critical issue of defining hypertext (cf. Nelson 1987, Kuhlen 1991, Foltz 1996, Huber 2002, Bublitz 2008). As usual with the descriptive efforts of this kind, the resulting picture remains rather diffuse. Usually, linguists embrace one of two frequent viewpoints on the phenomenon: hypertext is either framed as *electronic* form of text organization and navigation (cf. Bolter 2001) or as multi-sequential art of writing and reading paramount to both electronic and printed media (cf. Nelson 1987, Bucher 1998). One of these perspectives we could call *the restrictive view* and the other one *the continuity view*.

On the one hand, scholars advocating the restrictive view on hypertext discuss a constrained group of hypertext features to argue that hypertext is a strictly electronic phenomenon different from printed forms of hypertext. The restrictive view therefore proposes that there be a clear-cut division between printed and electronic forms of multi-linear writing (in order to emphasise a media-related surplus of the electronic communication).

Researchers which adhere to the continuity view, on the other hand, emphasise the pervasive similarities and continuities between what Ansel Suter (1995:11) calls printed *proto-hypertexts* and their new electronic counterparts.<sup>32</sup> Arguably, the first hypertext definitions which surfaced in linguistics adhered to the restrictive view, while more recent studies show a clear tendency towards stressing the continuity dimension of hypertext. In contrast to the restrictive view which stresses the electronic nature of hypertext features, the continuity view stresses the continuation of hypertext features from print-based to electronic documents. The duality of positions allows for a third view on the phenomenon. We shall call it *the integral view*, and it is the one advocated in this study. The integral view considers both the dominant socio-technological distinctions between printed and electronic forms of hypertext while at the same time eliciting the typological similarities which

---

<sup>32</sup> Printed predecessors of electronic hypertext are elsewhere referred to as *proto-hypertexts* (Ansel Suter 1995:11).

exist between them. It attempts to uncover the (implicit) intermedial dimension between text types of old and new media and also describes their media-specific residue.

These perspectives notwithstanding, there is generally agreement on the fact that a hypertext rests on an underlying network of documents, so-called *nodes* or *lexias*<sup>33</sup>, which are interconnected by some form of linkage. The result is an associative network of nodes which enables readers to choose their individual reader paths through the textual whole. Briefly, the multi-linear arrangement of nodes allows for multi-linear traversals.

Now, in most traditional print texts these receptive liberties are heavily constrained. For instance, we commonly go about reading classic prose by following a certain organizing principle. With this *principle of linearity* in mind, we intentionally start reading a book at the beginning, i.e. on the first page (possibly with the book's introduction or a preface) and then move on to the following chapters. Authors usually expect their readers end reading their book at the last pages which habitually consist of some sort of closure (preferably an epilogue or a bibliography).

It is important to note that book authors can certainly not force their readers to follow this rigid order of reception. Moreover, readers often escape reading documents in any 'preferred' order. Still, we might nevertheless argue that exceptions to the implicit convention simply underscore the fact readers are at least disposed to do so on a common basis. It is usually only when we travel to other (possibly Asian) countries which organised their documents in different ways that we understand how much we rely on textual conventions to understand documents. In fact, linear reading conventions (e.g. from left to right in the western hemisphere) have evolved to help us ascribe meaning to syntagmatic arrangements of signs. Readers usually expect authors to equally meet these conventions in some form or other. In composing their document, authors (in most cases) comply with this 'unsaid' convention. In the same vein, authors can readily assume their readers to read their document accordingly.

In a way, authors and readers thus base their activity on this very useful implicit principle or agreement. As a result, the textual arrangement of western documents, i.e. their internal order of chapters, their names and numbers, solidifies and perpetuates the principle of linearity. But let us change the focus and not only consider the creation and interpretation of books but the interpretation of its

---

<sup>33</sup> Roland Barthes defines *lexias* as "series of brief, contiguous fragments" (Barthes 1974:13).

components, i.e. sentences or utterances. We know that when hearing a series of utterances or a number of sentences, we understand the emerging linear segments of discourse to mirror the series of events denoted by them. This direct mapping of sequence of form and sequence of (denotative) meaning is often referred to as *diagrammatic iconicity* (Hiraga 1994). When we process contextually new verbal or visual information this technique enables us to apply meanings to verbal or visual syntagms.

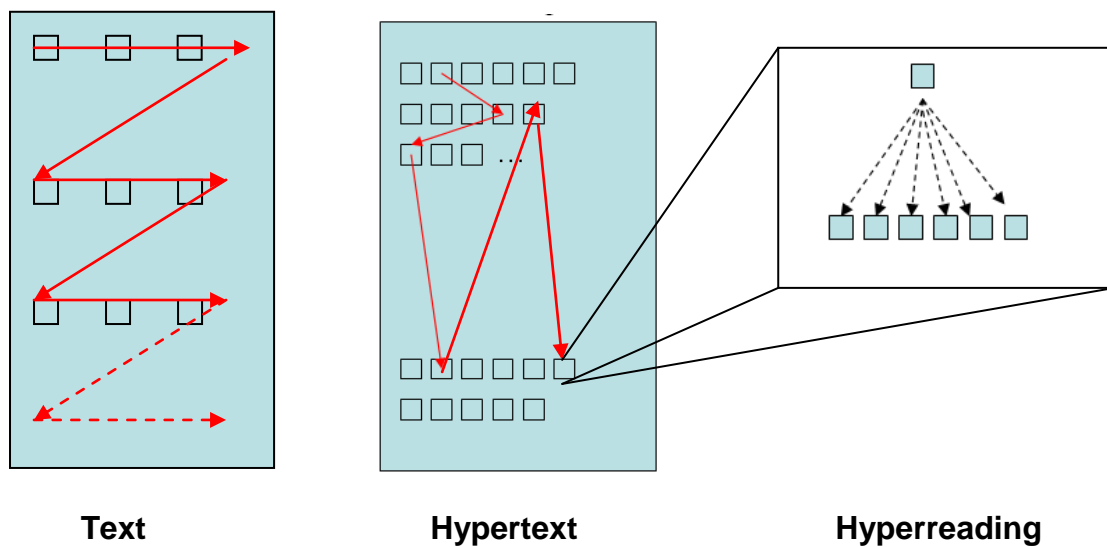


Figure 15: Prototypical Composition in Classic Linear Prose and in Multi-Linear Hypertext

However, the linearity principle does not imply that one must always read entire texts from beginning to end. In this context, we may thus distinguish two different concepts of linearity: *compositional linearity* and *perceptive linearity*. If readers follow the compositional linearity of a text, they choose to read an entire text in sequential order. In other words, readers are pursuing the author's intended 'line of thought' as it is believed to be prefigured by the linear arrangements of book chapters. Compositional linearity is only a regularity; a reading convention. Its practice is neither obligatory nor sufficient for text comprehension. Indeed, we frequently deviate from the deliberate habit of reading texts in linear order.<sup>34</sup> We are suspending and multiplying a suggestive linear text organisation. And when we do, we leap from one

<sup>34</sup> These deviations do not efface the concept of a linear reading altogether. Furthermore, they enhance the notion in the form of dispersed linearity. The reading process (here understood as cognitive perception) must necessarily remain linear.

paragraph to the next or from one note to the other. We browse and re-read text units, searching for comments or statements that are enclosed in the textual maze.

In fact, we constantly depart from sequence, citing things ahead and behind in the text. Phrases like “as we have already said” and “as we will see” are really implicit pointers to contents elsewhere in the sequence.  
(Nelson 1987: 1/17)

According to Nelson (1987) readers are used to perceive documents in an erratic, temporary fashion. Indeed, when there is little time, we tend to skip and scan texts looking for essential information, rather than indulge in slow reading. Some texts like dictionaries, telephone directories, shopping lists are meant to be scanned and skimmed. Therefore, they are composed in a fragmentary way in order to facilitate the quick-reading technique. They contain multiple self-contained text units to be read and connected in various directions. So, one could say that these texts essentially follow a principle of compositional multi-linearity rather than compositional linearity.<sup>35</sup> But how do we actually perceive rather than read these texts? Can we actually read in a multi-linear manner?

When we “depart from sequence” (Nelson 1987), what we do is actually a useful mixture of sequential and multi-linear reading. Reading is multi-linear in the individual goal-driven selection of reading paths (*navigation*) but it is sequential in that the perception of visual signs along the chosen path (*reading*) is linear. In other words, hypertext reading (or hyperreading) is spatially multi-linear and temporally linear (see figure 8).

As we can see hypertext adheres to compositional multi-linearity on a textual level while it simultaneously endorses sequential linearity on a clausal level. Any deviation from the compositional linearity must lead to a *re-cognition* of a new self-selected linear sequence. Surprisingly, the multi-linearity of hypertext therefore does not dispose with the concept of linear reading altogether. Linearity is not abolished, rather it is dispersed.

Accordingly, compositional multi-linearity can be subdivided into two constitutive processes: hypertext creation and the hypertext construal, i.e. *hyperwriting* and *hyperreading*. Hyperwriting amounts to the purpose-driven

---

<sup>35</sup> We should note at this point that it does not simply amount to what others have called non-linearity (Landow 1997:61): “No, hypertext is not about lack of linearity. For me, hypertext is about the necessary combination of nonsequential and linear. There is never a lack or complete absence of linearity.” (Carolyn Guyer, personal communication during an on-line conference at PMC-Moo, October 1993.)

compositional alignment of self-contained hypertext units towards a network of nodes. Hyperreading refers to the hermeneutic exigencies which are involved in the planning, selection and reading of the hypertext. Both hyperwriting and hyperreading, therefore, include the dimension of perceptive linearity as a minimal structuring device. Productive and receptive linearity exists within the complex of larger text components of the hypertext (see figure 16). It is constantly explored, enhanced and redirected in discourse by both hyperwriters and hyperreaders:

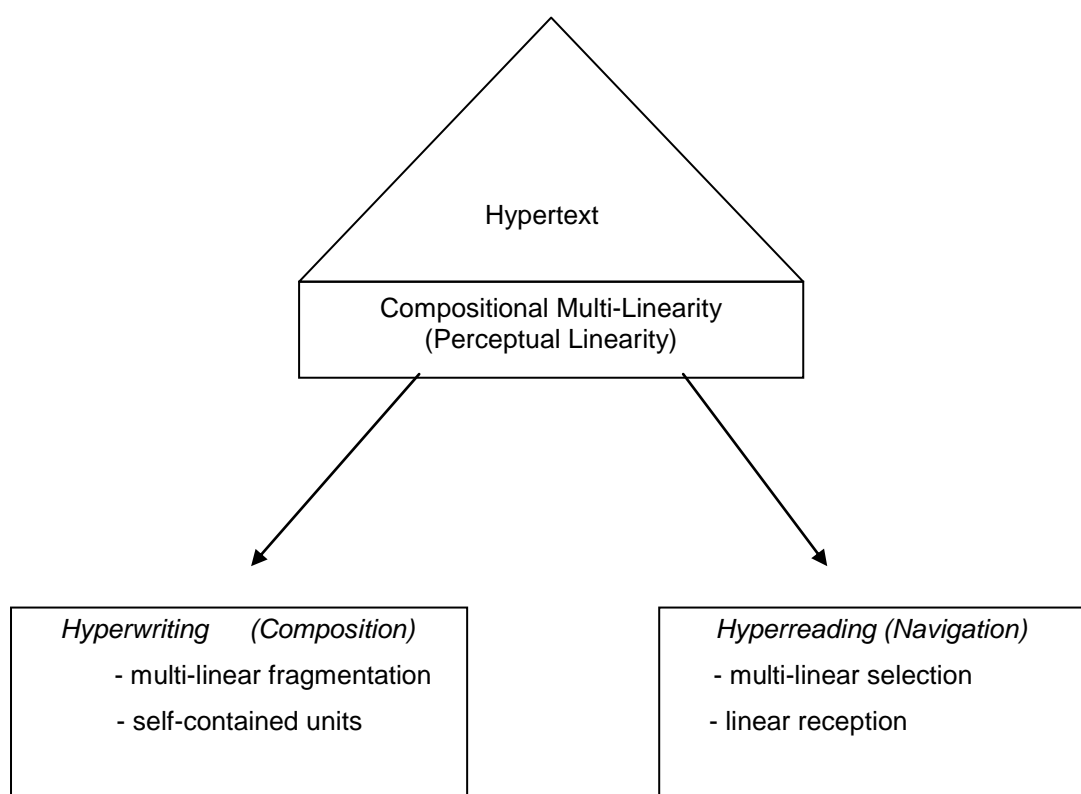


Figure 16: Hypertext Hermeneutics

We may now use these preliminary reflections on the nature of hypertext as a springboard to exploring the medial dimension of compositional multi-linearity. As we have seen, the concept is present in both printed and new media. We might spot a range of multi-linear text formations in various print-based text types, some of which actually possess quite extensive link typologies. *Gamebooks*, a specific type of multi-linear fiction, can serve as a paradigmatic example. Each text passage of a gamebook connects to other text units. As a reader, one has a number of narrative choices which materialize in form of specific “points of departure” at the end of a particular text unit, e.g. ‘to enter the dark room go to page 24’. Such text links are

consequently essential indicators of future reading paths. They both organize and anticipate the story path which lies ahead.<sup>36</sup> Skipping to a particular page of their choosing, readers may thus co-determine the development and outcome of their story. The web of interrelated text units permits readers to select their individual narrative continuation from a number of possible prefigured storylines (figure17).

Beside gamebooks, a similar web of links appears in other printed text genres, such as dictionaries, phone books, etc. Here, links do not progress various narrative sequences but serve to refer readers to selective pieces of linguistic or encyclopedic information.

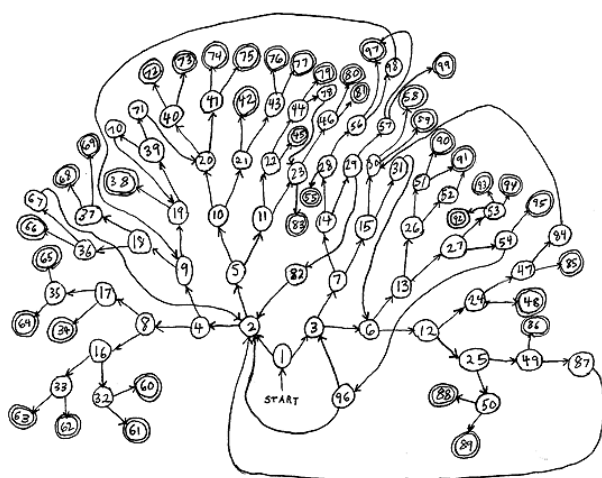


Figure 17: Conceptual Sketch of Gamebook Reading Paths  
(<http://Internet.gamebooks.org/canedit.htm>)

Although both genres provide an internal system which organizes the texts, e.g. the alphabetical order of entries in dictionaries, their internal composition is not devised for readers who wish to peruse content in linear order. Moreover, readers are encouraged to browse text data selectively, according to their personal interests. However, both gamebooks and dictionaries differ with respect to their presentation, placement, number and type of linkage. It therefore seems beneficial to scrutinize links in more detail to ascertain some key differences between various types of hypertext. What is more, it seems likely that differences between printed and electronic hypertext are motivated by their respective differences in the form and

<sup>36</sup> "Links are intrinsic to documents, and have been for millennia. A link is simply a connection between parts of the text or other material." (Nelson 1987:2/23)

usage of their links. In the following, links used in traditional (printed) hypertexts shall thus be named *text links*, and I shall refer to links used in computer-mediated environments with the term *hyperlinks*. Let us therefore discuss the key link features which help us to distinguish “old” analogue and/or “new” digital forms of hypertext.

### **3.6 Across Media – From Analogue to Digital Hypertext**

There are, of course, multifarious similarities between text links and hyperlinks. For instance, both types connect multi-linear text parts or units and both have a certain point of departure. Both also point toward a certain target referent which they relate to a source paragraph. They can be positioned adjacent to the continuous text flow or indeed be located within the primary text itself. Text links are, however, also different from hyperlinks. Let us look at these in more detail.

Let us start with conventional form. On the Internet, electronic hyperlinks are presented in different forms, lengths, sizes and colours. They do not follow any major formal restrictions except that they should connect two entities and be electronically operable. As a result, some hyperlinks are not detectable as such at all. Text links, on the contrary, rely on specific typographic conventions. In dictionaries, for example, explicatory tables and lists of indices remind us of the complexity of these formal denominations. As it turns out, links are not only constrained by a consistent typographic characterization but are also dispatched to specific parts of the page. These compositional traits have been established over centuries and they answer to concurrent medial and generic affordances.

In this sense, classic text links have gone a long way toward their own generic naturalization. Hypertext is still on the verge of establishing such decisive generic traits. Accordingly, hyperlinks have yet only answered to temporary, often inconsistent design trends. There is no discernable norm which clearly regulates the formalization of hyperlinks with respect to different Internet text forms. Hence, one central characteristic of electronic hypertext is its lack of stable conventions for the formal and compositional display and placement of hyperlinks.

Another contrast between text links and hyperlinks regards their medial representation. Classic text links occur in printed form. *Ergo*, they materialize on an underlying medium, usually paper. Traditional text links thus leave physical traces on this medium, for instance ink on a piece of paper. Hyperlinks do not leave traces on screens but on hard drives. They connect transcoded documents as decomposed



digital chunks of data (bits and bytes).<sup>37</sup> As a result, new pages appear on screen while others recede or vanish from sight. Hence – visually speaking – hyperlinks are “immaterial” entities. In contrast, text links are physically present and thus more stable than hyperlinks.

Text links are fixed to some preordained location which makes them difficult to efface. On computers, however, the visualization and storage of data are two separate processes which pertain to two material devices, i.e. the computer screen and the hard drive.<sup>38</sup> This media-related difference between printed and electronic text links has repercussions. The separation of data storage and visualization allows computers to save and store more data than a book. It equally enables readers of electronic hypertext to easily access data quicker than in printed hypertexts:

In a book, one had to read from left to right (or right to left, or up to down, according to different cultures) in a linear way. One could obviously skip through the pages, one - once arrived at page 300 - could go back to check or re-read something at page 10 - but this implied a labour, I mean, a physical labour. On the contrary a [electronic] hypertext is a multidimensional network in which every point or node can be potentially connected with any other node. (Eco 1996:6)

Hence books simply always have a rather limited storage capacity. A search for specific paragraphs in books (or indeed for books in the vast aisles of a library) is usually cumbersome. Another key advantage of the electronic hypertext is therefore its ability to reduce and conflate textual data individually.

We can already surmise that computers support hyperreading more effectively than books because their digital processing facilitates the multi-linear reading process (cf. Freisler 1994:20). While books make ample use of intertextual references or use links to index related sources of information (e.g. bibliography), readers will still need to search for these sources individually, some of which might even be difficult to obtain. In short, the book as a medium simply cannot reach the complexity of a computer or computer network, as it were:

---

<sup>37</sup> “In new media lingo, to *transcode* something means to translate it into another format of expression. The computerization of culture gradually accomplishes [...] transcodings in relation to all cultural categories and concepts.” (Manovich 2001:47)

<sup>38</sup> Friedrich Kittler observed that “the transition from writing to the technical media [was] indeed [...] a decoupling of communication and information” (Kittler 1996: no pagination). Similarly, Baldry & Thibault (2005) conclude that “the material support and the tracings on the [surface of a book] are *hard-coupled*” and “fixed”. On the contrary, “the web page is characterised by the *soft-coupling* of material support [...] and data (digital bytes)” (Baldry & Thibault 2006:109). Loosening the medial fixation of data facilitates the processing of multi-linear information.

[hypertext is] written or pictorial material interconnected in such a complex way that it could not be conveniently be presented or represented on paper.  
(Nelson 1965:96)

Still, one has to acknowledge the fact that all these technological advantages of hyperlinks come at a price. Printed texts have merits of their own. For example, hypertexts lack the material permanence and tactile quality of their printed counterparts. Some would say books are lighter and therefore easier to travel with or that they do not strain the eyes (due to higher resolutions).

This line of argumentation is often referred to as the “bathtub argument.” It states that the applicability of electronic books is greatly limited by the physical constraints induced by computer hardware. Computers, according to this view, are allegedly restrained by its use to few working environments only. In other words, computers are said to be too large, bulky or even dangerous for its use in everyday places like bath tubs, planes, cars, camp grounds, etc.<sup>39</sup>

We can now produce a number of key oppositions between hyperlinks and text links (see table 6). Although these may not be complementary but gradable differences, they allow us to deduce a number of features which are either more prototypical for printed or for electronic forms of hypertext (see table 6).

<i>Characteristics</i>	<i>Digital Hypertext (Comp.)</i>	<i>Analogue Hypertext (Print)</i>
<b>Connection</b>	Hyperlink	Text Link
<b>Connectivity</b>	Automatic	Manual
<b>Accessibility</b>	High	Low
<b>Rapidity</b>	High	Low
<b>Dynamics</b>	High	Low
<b>Complexity (Size)</b>	Medium – High	Low – Medium
<b>Conventionality</b>	Low – Medium	High
<b>Stability / Fixation</b>	Low – Medium	High
<b>Visual Resolution</b>	Medium	High

Table 6: Printed (Analogue) vs. Electronic (Digital) Linkage

<sup>39</sup> Literary scholar Jay D. Bolter (2001) is known to have criticized the test by claiming it did not distinguish text from media. He pointed out that computer hardware was already being designed to suit the purposes of reading in various human environments (e.g. water-resistant floatable electronic books, light and portable reading devices, e.g. amazon's *kindle*®).

So far, I have only considered general characteristics of the hyperlink in order to distinguish printed from electronic hypertext. We have seen that hyperlinks make for more flexible, rapid and easy access to and retrieval and storage of electronic data. On computers, hyperlinks appear in different numbers and are allocated to various locations of the web page. Consequently, there are different types of electronic hypertexts. Some of them might have a very limited number of hyperlinks which connect text units in linear succession, others may be more complex. We might even question whether some electronic texts are hypertexts at all. Is it not more convincing to state that electronic texts, whose parts are successively linked, are merely electronic imitations of classic linear prose? How do we distinguish these linear entities from “actual” multi-linear hypertext?

Storrer (2002) offers one viable solution for this problem. She calls linear electronic documents *e(lectronic) documents* and compares them to proper *hypertexts*. Storrer’s hypertexts contain more hyperlinks than e-documents. Hyperlinks are not only located *hors-text*, as it were, but also to be found within single, self-contained text units.<sup>40</sup> As a result, lexemes, phrases, sentences and paragraphs may be linked to various target areas in and across hypertexts. The individual scope of hyperlinks thereby differs extensively. In this sense, *hyperdocuments* are what Storrer (2002) calls hypertexts as long as they share one discernible theme, perspective or text function. Other hypertexts, which are thematically too complex to be treated as a stack of thematically or functionally related documents, Storrer names *hyperwebs*.

The Internet itself is at the same time composed of e-documents and hyperdocuments. In some way, the Internet itself represents the largest of all possible hyperwebs. It is itself, however, subject to an indefinite mass of additional subsidiary hyperwebs. In terms of linkage, both hyperdocuments and hyperwebs consist of large numbers of hyperlinks located either inside or outside single text units. In regard to the composition (layout and design) and scope (size of source and target unit) of their individual hyperlinks, hyperdocuments and hyperwebs are only distinguishable by degree.

Still, the tripartite concept of e-texts, hyperdocuments and hyperwebs provides a common ground for the analysis of different types of hypertext which are opposed

---

<sup>40</sup> Hoey (2005:74) classifies these text units as *colony texts* defined as text segments “whose random excision will not impair or render intelligible the entire text.”

in regard to features like link quantity, theme or function. On the Internet, we can detect a gradual move from linear to multi-linear text arrangements (and this fact is mirrored by the evolution of Weblogs as illuminated in chapter 2). Such multi-linear text arrangements are native to weblog texts. Entries and comments are not only interrelated by their spatial proximity on screen but, more crucially perhaps, by hyperlinks. We will see at a later stage of this study that hyperlinks have a considerable role to play in the acquisition and negotiation of shared knowledge between authors and users.

Arguably, multi-linear text compositions aim to synchronize the textual representation of information to the representation of knowledge in the human mind. Indeed, the impetus for writing multi-linear texts was historically preceded by people's urge to convey more effectively. It has been assumed that hyperwriting is deeply rooted in what in the way we think. But how can one conceive of hypertext cognition? How does it tie in with our concept of hyperwriting and hyperreading? These questions shall be approached next.

### **3.7 Across the Mind – Hypertext and Cognition**

At the root of hypertext theory is the firm belief of its founding fathers (Nelson 1965, Bush 1945) that the way we communicate should mirror the way we think. This idea was first advanced to the scientific community in 1945, in an article called "As we may think" written by Vannevar Bush. The author recognizes a growing need in western societies to invent machines which are capable of enhancing human abilities to cope with a "growing mountain of research" (Bush 1945:4). To this end, he suggests the construction of a mechanical device called 'the memory extender' (henceforth: *memex*). A memex was to store a vast amount of information on micro fiche accessible through specific terminals. Terminals were conceived as some form of visual 'desktop'. A number of levers were attached to the terminal interface, and users were believed to control them to access and traverse the micro films.<sup>41</sup> Users could now traverse the documents stored on micro film in various ways and directions. The single pursuit of users' reading path was meant to be performed some sort of hyperlinks which connected the documents to each other in multi-linear ways.

---

<sup>41</sup> Lemke (2002) decomposes the notion of hypertext reading into two subsidiary concepts: *trajectory* and *traversal*. He calls trajectory "sequence[s] of presented signs" composed by the hypertext author and distinguishes it from traversal defined as the "interpreted meaning experience created by making meaning along the trajectory." (Lemke 2002: Appendix)

Reader's individual link choices left mechanical traces by perusing the data which was stored on the memex in a systematic fashion. The aim was to store the traversals to be able to retrace reading paths at later stages upon individual inquiry. With the help of the memex, Bush thought to have created a device that could map and retain users' cognition by storing reading trails across documents. In other words, thoughts were meant to be reified on documents (or micro films):

The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade, items are not fully permanent, and memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature.  
(Bush 1945:45)

Interestingly, Marshall McLuhan (1964) would argue few decades later that media are extensions of men. In this light, Bush's invention can be remodelled as a medial prosthesis to the human mind. In fact, the memex was expected to transduct human thought into lasting form (e.g. book, papyrus, stone) but retaining its innate multi-linear structure. One sought to conserve the rhizomatic form of cognition in writing and linear ways of writing came to be regarded more as a formal imposition on cognition rather than an asset for logical argumentation.<sup>42</sup>

In the Phaedrus dialogues, Socrates dismisses the written word because of its alleged propensity to force the erratic structure of cognitive processes into an unnatural linear order. The fixity of writing, he feared, would deprive communicants of their ad hoc right to negotiate and clarify discourse meaning and function.<sup>43</sup> Indeed, written words principally lack any physical connection to the human mind. Words or rather lexemes are deliberately withdrawn from the actualization of the communicative event. They only act as indices or signposts for possible interpretations, rather than bearing actual meaning themselves. In other words, words are the visible footsteps in the linguistic snow of texts. We should not fail to

---

<sup>42</sup> The metaphor of hypertext as a root or *rhizome* is used here to address the unstable dynamics of hypertexts to grow into larger, multi-sequential networks of data (cf. Deleuze & Guatarri 2007).

<sup>43</sup> Socrates is likewise said to have lamented the fallacious state of written communication. According to Plato's writings, he apparently feared that the "forgetfulness" that the written word allegedly provoked "in the learners' souls, because they will not use their memories; they will trust to the external written characters and not remember of themselves."

(<http://Internet.gutenberg.org/etext/1636>, 15.09.2009)

recognize the internal complexity of discourse meaning and give in to the fallacy of proclaiming a direct relation between words and meanings:<sup>44</sup>

There is a kind of illusion in the world we live in that communication is something that happens all the time [...]. Actually, communication is an exceedingly difficult activity. In the sense of a mere point-to-point correspondence between what is said, done, and thought and felt between people – this is the rarest thing in the world - If there is the slightest tangential area of touch, agreement, and so among people, that [sic!] is communication in a big way.

(McLuhan 1964, cited in Cavell 2003:5)

Nonetheless, there is, of course, a contrast between the cognitive structure of our minds (which is organized multi-sequentially) and classic written text (which we perceive to be ordered in various linear parts). People make use of diverse techniques and methods to bridge this disparity. Let us consider the medium *voice*, for example. The topical and argumentative linearity of oral discourse can be disrupted by interlocutors. They can engage in digressions, interruptions, topic shifts, queries, explanations, repairs, questions, etc.. Written prose, on the contrary, denies readers this kind of control over content, readers cannot engage verbally in the discourse at hand. In accordance with Socrates' loss of immediacy, writers defer the actual act of communication by detaching themselves from the interaction. In print media, readers no longer negotiate meanings directly with authors. Print text is petrified, the ensuing (inter-)action dismantled and a state reached in which hermeneutic reflections can only be validated with great difficulty.<sup>45</sup> Oral discourse is temporally linear but conceptually multi-linear because of its interactive potential which allows interlocutors to momentarily put their discursive orientation at disposition. In face-to-face conversations, topics can be skipped, shifted, enhanced or stopped by various interlocutors. In contrast, most written discourse is organized temporally and conceptually in a linear fashion; here the same dynamic imposition on the organization of discourse is largely unattainable.<sup>46</sup> Therefore, in oral discourse, negotiating meaning means "fixing the floating chain of signifieds"

---

<sup>44</sup> Montaigne claimed that "all memory [was] consigned to the characters printed on the page." His books served him to regain his memory. Still, it obviously did not produce the memory by itself because it still needed the activation of its reader.

<sup>45</sup> This, of course, does not strictly apply to quasi-synchronic forms of electronic written communication, e.g. chats.

<sup>46</sup> Even electronic hypertexts are not in any way different to classic print texts in this respect. They may allow users to traverse texts in different directions but users cannot impose on the content itself. Weblogs, however, are what Ong (2003/1988) has called a *second orality*, i.e. an augmented kind of participation in written form.

(Barthes 2001/1977:39). In written texts, the inaccessible author gives birth to an almost “infinite deferment of the signified” (Barthes 2001/1977:158) but controls the organization and sequence of the discourse. The interactive potential of oral discourse therefore lends itself particularly well to the multi-linear organization of content. This manifests itself most prominently in the communicative duality of turn-taking.

The process-related analogy between discourse organization and cognitive information processing recurs in the form of Socrates’ idea of immediate hermeneutic approximation of conversational meaning. The idea in classic philosophy has thus come to fruition only through printed, not through oral form. While our culture has been transformed considerably by the linear conception of the written words, people have not ceased to question the totality of compositional linearity. Most notably, Ludwig Wittgenstein rigorously defied the idea of linear writing. In the preface to his *Philosophical Investigations*, he surrenders to the inevitability of forcing his thoughts into the rigid schemes of written word:

But it seemed to me essential that in the book the thoughts should proceed from one subject to another in a natural, smooth sequence. After several unsuccessful attempts to weld my results together into such a whole, I realized that I should never succeed. The best that I could write would never be more than philosophical remarks; my thoughts soon grew feeble if I tried to force them along a single track against their natural inclination. – And this was, of course, connected with the very nature of the investigation. For it compels us to travel criss-cross in every direction over a wide field of thought.

(Wittgenstein 2004/1953:7)

In taking with Wittgenstein’s dilemma, the philosopher David Kolb (1994) reviews the same dilemma of forging one’s mind into the compositional restraints of the written space. But to Kolb (1994), electronic hypertext figures as an excellent candidate for retaining both the argumentative integrity of paper and the multi-linear form of thoughts. Kolb maintains that philosophical writings have to retain a primary line of progression for argumentative purposes. In other words, writing – or in Kolb’s case: philosophical argumentation – must in principle adhere to a line of reasoning. He acknowledges, however, that this line might itself be expressed and experienced along diverse multi-linear trajectories. Therefore, he considers the personal selectivity of hyperreading to be a loophole to Wittgenstein’s ‘linearity conundrum’. Indeed, he shows that hypertext endows linear writings with contextual “supplements which it both desires and rejects” (Kolb 1994). In traditional texts, Kolb holds, text links fulfil

the function of aligning these supplements, e.g. “marginalia in medieval manuscripts, parallel columns of text and material in parentheses or footnotes” (Kolb 1994:no pagination). In fact, connectives and asides have already been integral components of printed texts and served the purpose of providing a contextual sphere to the continuous flow of text. Furthermore, digital hypertext builds on these traditions but embraces and multiplies the accessible supplements. Electronic hypertext facilitates access to the complex interrelations of ideas, thoughts and argumentations. In this sense, we might posit that Wittgenstein and Kolb hold similar views. On a different level, we might spot similarities between the authors’ frame of mind and Nelson’s original vision of hypertext:

The structures of ideas are never sequential. They tie together every which way. And when we write, we are always trying to tie things together in non-sequential ways. [...] People keep pretending they can make things hierarchical, categorizable and sequential when they can't. [...] In an important sense there are no 'subjects' at all; there is only knowledge, since the cross-connections among the myriad topics of this world cannot be divided up neatly. Hypertext at last offers the possibility of representing and exploring it all without carving it up destructively.  
(Nelson 1983:29)

It follows that hypertext does not only consist of a stable network of text units (“a cloud of written commentary”, Kolb 1994) but likewise consists of a range of associative (not hierarchical) relations which connect them. Therefore, hypertext manages to illustrate the rich layers of association which surround an ongoing discourse more conveniently than printed documents. In addition, hypertext’s disregard for textual closure puts an end to strict argumentative unity. More specifically, hypertext users exclusively determine whether some text has come to an end. Additionally, the access of hypertexts via different reading paths makes possible superficial and thorough readings of texts. Users may browse through text units in a straightforward or elaborate way. They might activate or neglect hyperlinks and these links can lead to associative or unrelated text units, which hold additional but possibly also outdated information. Likewise, hypertext fragmentation may, but does not automatically, amount to an unstructured, incoherent multimodal patchwork of data. Moreover, hypertexts offer technological tools, e.g. hyperlinks, menus, search engines, site maps, etc., which enable authors to visualize specific text segments. Websites are embedded in an overarching network of different, albeit related connections. The difference between indecisive groupings of text units and cohesive



segments of multi-linear text is partially dependent on the ability of hypertext authors to structure, segment and tag the content of their text units effectively. It does, however, equally rely on the users' competence to suitably integrate electronic and textual ties among hypertext units to create a larger coherent whole. Therefore, let us now investigate how hypertext authors and their readers approach hypertexts.

### 3.8 Across Space – Knowledge, Discourse and Participation

Traditionally, text creation (just as text interpretation) is a complex and multi-levelled process, especially in hypertextual forms of writing. It demands a number of technical and practical skills from their creator (or indeed creators). In traditional mass media, different individuals were involved in the production of single texts, each playing a specific indispensable role in the overall process. While typically sharing a common objective, these “actors” assumed different levels of responsibility and authority with regard to the actual outcome of the creative process, i.e. the textual artefact. If a journalist, for instance, authors an article, he does not control the theme, style of the subject matter solely by himself. Rather his task is guided and determined beforehand by various collaborators, like the editorial board or his editor in chief. The latter determines the general purpose, style and theme of the article. We might call this person the *principal* (cf. Goffman 1982:145f.).<sup>47</sup> He initiates the actual process of text production which is thereafter enforced by the journalist who we may call the *author* (Levinson 1988). The author draws on directions provided by the principal to design his article. The process of writing the article is finally rounded off by a third type of actor, the *printers*, who position the text in line with other articles on the newspaper page. In sampling the various articles into suitable format, they will finalize the papers' individual appeal. The printers are responsible for the final text composition of the article. They establish a recurrent text design in which the semiotic content of the article is entrenched. Finally, the article must be delivered to preordained social venues, e.g. newsstands, in order for it to “inter-act” with the common consumer, i.e. the reader. The person engaged in this final step of the production process (whoever he or she might be) is called the *animator* or *transmitter* of the message. All of these productive steps and roles are undergone (in one way or another) in classic mass media text production. Traditionally, the phases are realized

---

<sup>47</sup> Goffman (1982:145) describes as *principal* a person who imposes on the writing process without actually participating in its material genesis.

by different people in a time-consuming chain of individual, yet interrelated processes.

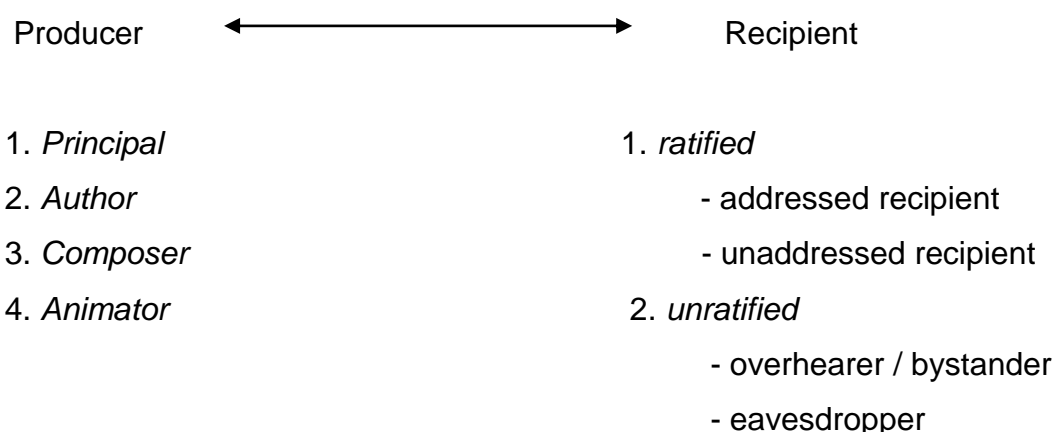
In contrast, blog authors are expected to perform all of these tasks at once. She will choose a design template from an online blogging service and personalize it with a distinct text design (composer). She decides when, how and why to update her weblog's content (principal) and writes – and links – the actual content of her posts (author). Although she may not be the animator of her texts, a blogger may engage in many interactive practices surrounding the posting of the article (which itself carries some sort of animator quality). She could, for instance, engage in comment areas, promote her weblog posts in other online fora, print her weblog post or link it to some other online resource. Weblogs thus represent a re-unification of Goffman's *production formats*.

On the receiving end of discourse, the picture looks quite different. In Weblogs, there is no concurrence of recipient roles such as there is with production formats. Some users can gain access to a weblog some of which may be intentionally addressed by the blogger, others may not. Again, we may draw on Goffman (1982:146) who distinguishes between *ratified* and *non-ratified recipients*. While ratified recipients may be explicitly addressed or unaddressed individuals, bloggers will not intend to direct their attention to unratified readers. According to Levinson (1988), unratified hearers are either *overhearers* or *eavesdroppers*. The first term *overhearer* (in weblogs: *overseer* or *bystander*) refers to individuals who might just happen to perceive a message coincidentally, for example a passer-by who overhears snippets of an ongoing conversation.

*Eavesdroppers*, on the other hand, are bystanders which remain unnoticed by the speaker-writer. They deliberately act on perceiving some form of audible or readable information without the prior knowledge or even permission by the respective author. In Weblogs, authors usually call upon a certain group of people in their weblog texts or weblog rolls. This group of people can be considered the regular ratified audience of the weblog. Bloggers, however, operate on the knowledge that a large mass of unratified weblog users may easily access and peruse their entry and comment sections. Actually, in most cases, bloggers will personally try to evoke the interest of anonymous readers in order to appeal to a larger, yet unratified audience. In order to achieve this goal, weblog users may assume ratified status; such is what happens when bloggers directly engage in comment sections to interact with “new”

commentators on their site. After unratified users have voiced their opinions and revealed their names in comment sections, the weblog author is free to raise their status or ratification by replying. Should a previously unratified weblog reader emerge in weblog comments, a blogger may then either choose to ignore (so as to accord him bystander status) or address the user directly (so as to ratify the reader status). Indeed, in weblogs communication roles are not only dispersed across a myriad of potential users, the status of users themselves is equally transient and open to considerable change. Similar to face-to-face conversations, where bystanders can become ratified listeners; in weblogs, unknown users can turn into ratified readers by taking discursive shape in comment sections.

### Levinson's Participation Roles



### Decomposition of Participation Roles in Weblogs

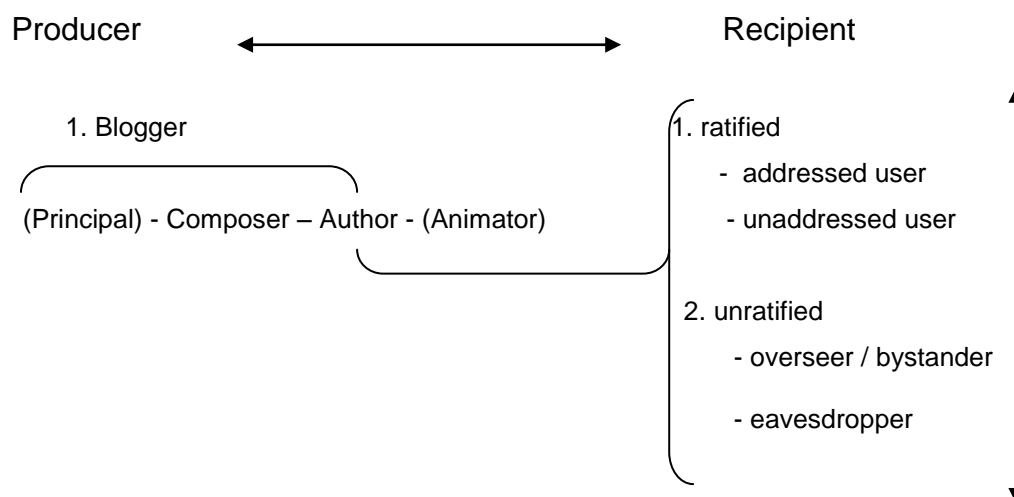


Figure 18: The Unification of Participation Roles in Weblogs

In compliance with the diverging status of weblog users (ratified vs. ungratified audience), a blogger will compose his text with a view to the previous knowledge of a particular (narrow or broad) target audience. Likewise, users will thus assess the coherence of weblog texts on the basis of their prior knowledge which they assume is shared between interlocutors, i.e. the common ground. Let us therefore explore in more detail the knowledge domains from which bloggers and users draw to construe common ground in ongoing discourse. Common ground, i.e. shared knowledge, is built up from knowledge which Bublitz (2006:368) allocates to three cognitive domains: linguistic knowledge, immediate knowledge and declarative knowledge.<sup>48</sup>

According to Bublitz (2006:372), linguistic knowledge is twofold. It is information *about* language as well as information on *how to use* language in various contexts. As such, *linguistic knowledge* refers to a number of linguistic domains ranging from the identification of phonemes and variants to the rather complex task of recognizing discursive moves.

*Immediate knowledge* pertains to knowledge about the situational parameters (components of the discourse situation) which Bazzanella (2002:241) describes as being composed of a) discourse participants (their age, [gender], origin, *status*/social role, and shared knowledge/beliefs), b) the physical (i.e. time and space) and cultural setting, c) the types and goals of the interaction, d) instruments used for the purpose of communicating (channel, dialect, etc.) and e) the preceding or following verbal text (*co-text*). Finally, *declarative knowledge* connects discourse to its social, cultural and historical context of use. Declarative knowledge relates to our knowledge of facts and events in our lives as well as the prototypical settings to which we subscribe in everyday interactions.

It seems evident that the three categories share an inherently analytical character and are as such not as straightforward as the previous introduction might suggest. It is likely that all three spheres of knowledge are interlaced, at least to some extent. Still, the three-fold division of knowledge is useful to us because it allows a first tentative exploration of the cognitive field.

---

<sup>48</sup> In addition to the three knowledge types presented here, Bublitz (2006:369) also mentions empathetic and cognitive skills which are also important prerequisites for the construal of coherence. These skills are relevant to inferring, accessing, combining and blending various cognitive domains. As such, I consider them as cognitive skills or techniques and not as basic types of knowledge *per se*.

The vast domain of general, declarative knowledge can be further subdivided into two categories: *episodic* and *semantic memory* (cf. Tulving 1983:28) both of which can refer to knowledge on facts and/or events. Episodic memory revives our individual, personal experiences which we share exclusively with only a limited group of adepts. In contrast, semantic memory is assumed to be generally available to most members of a language community. In this sense, episodic memory is more socially-constrained while semantic memory is less-limited within a community of speakers.

Hence, declarative knowledge consists of both semantic and episodic memory, as facts or events might be shared by limited groups or larger parts of a culture. Facts and events differ to the extent that facts are structured and stable entities, while events are time-bound and dynamic. In this light, declarative knowledge consists of either *structural knowledge* which concerns the composition of entities (facts and topics) or *procedural knowledge* which refers to events in prototypical situations (sequences and procedures).

The latter procedural kind of knowledge thus ties in with the former structural one to build cognitive “packages”, construing stereotypical cultural patterns. These patterns systematize and guide our individual behaviour in social groups and forms of interaction. They provide the cognitive background for our discursive habits, formal expectations and cultural stereotypes by which we conceptualize the world around us. As a matter of cause, we are now able to define the following cognitive domains and express their internal structure:

<p><b>Linguistic Knowledge</b> – Semantic – Structural and Procedural</p> <p><b>Immediate Knowledge</b> – Episodic – Structural and Procedural</p> <p><b>Declarative Knowledge</b> – Episodic or Semantic and Structural &amp; Procedural</p>
---

In 1925, the French sociologist Maurice Halbwachs introduced the concept of *collective memory* in an effort to portray the temporal dimension of semantic memory. Interestingly, Halbwachs (2006/1925) did not describe cognition as a subjective condition or activity but rather described thinking in its underlying interactive dimension. For Halbwachs, the memory was not self-contained and subject to deliberate retrieval of information (e.g. the *art of memory*). Moreover, he figured that the memory was shaped through social interaction. Indeed, Halbwachs (2006/1925) claimed that memory is only conceivable through our cognitive interaction with

others. For him, memories emerged through a constant interplay of personal and communal experience.

Although the act of memorizing is thus individual in its information retrieval, it is intersubjective in its construal of facts and events:

To be sure, everyone has a capacity for memory [*mémoire*] that is unlike that of anyone else, given the variety of temperaments and life circumstances. But individual memory is nevertheless a part or an aspect of group memory, since each impression and each fact, even if it apparently concerns a particular person exclusively, leaves a lasting memory only to the extent that one has thought it over – to the extent that it is connected with the thoughts that come to us from the social milieu.

(Halbwachs 2006/1925:53)

Following this line of thought, Halbwachs (2006/1925) created his concept of *collective memory*. Through the act of remembering, Halbwachs explains, the individual reinforces, enhances, focalizes or neglects certain sets of values which are entrenched in the way societies have come to recall their history. The sharing of a consistent number of memories and values toward one's past means adhering to collective memory. Halbwachs suggests that collective memories are central to the coordination and attribution of human image and identity.

Furthermore, personal knowledge turns into collective knowledge through the process of communication. Just as personal knowledge is shaped and transformed through its adherence to different collective memories. The idea of a gradual participation in various groups of knowledge (collective memories) appears useful in the context of understanding documents like weblogs. Weblog users understand entries to the extent they share a certain common ground with the blogger. What we have termed common ground now aligns with user's access to collective memories similar to the ones of the blogger. Such a cognitive overlap between writer and users might be necessary to "fill in the blanks" left by the pragmatic insufficiencies of the respective text unit (ergo contextualize it sufficiently to construe a coherent text). Otherwise, diverging participation in these memory collectives can be expected to yield different sets of knowledge which can be expected to lay the ground for a number of text-based misunderstandings.

Halbwachs' idea of collective memories thus perfectly illustrates how knowledge itself nurtures social forms and norms of societies when used in interaction. In remembering, we cooperatively shape these memory spheres but equally absorb their values as part of our own identity. There is, however, another

crucial insight into the nature of declarative knowledge which we may gain from Halbwachs' theory. It is the fact that memory builds up gradually. It is dependent on time. First, the memory yields temporary forms of episodic memory which are then selectively filtered into more abstract facts and procedures stored in the semantic memory. Episodic memory thus represents a primary stage in the gradual accrual of semantic memory. So, there is personal "eye-witness" knowledge which we share with a restricted group of individuals for a limited period of time (*episodic memory*) but there is also a biographical knowledge based on personal experience gathered throughout our lives.

Finally, there is knowledge which is not based on eye-witness accounts at all but which we derive from specific media or artefacts, like history books. This knowledge is highly formalized and depends on ritualized, institutional acts of remembering, for instance, as a result of celebrating a national holiday. Based on Halbwachs' idea of collective memories, Jan Assmann (1992) developed a useful threefold distinction of memory types. Memory which is based on someone's personal experience is called *communicative memory*. This type of knowledge is temporally located between episodic memory and what Assmann (1992) calls *cultural memory*. As opposed to the communicative memory, the cultural memory exceeds the life span of a living individual, i.e. it does not principally rely on individual eye-witnesses. It is rather a socially condensed form of memory which we internalize through ritualistic cultural actions (celebrations, prayers, etc.). Whereas episodic memory is constrained to limited time periods (present - ca. 20 years), communicative memory entails self-witnessed quasi-biographical knowledge, which has been either collected by personal experience or gathered through face-to-face interaction (ca. 80-100 years). Cultural memory exceeds the life range of approximately 100 years. As such, it relates to memories passed down from generation to generation.

What Assmann (1992) calls cultural memory is accessible through conventional mnemonic rituals which involve the use of specific cultural artefacts. These objects have over generations acquired specific symbolic meaning in a given culture or community. They may be documents like photos, videos, texts but also monuments, places or particular sorts of food (e.g. the famous turkey at a

Thanksgiving dinner).<sup>49</sup> Assmann (1988:129) claims that “communicative memory is characterized by its proximity to the everyday, cultural memory is characterized by its distance to the everyday”. Hence, cultural memory is a collective kind of memory evoked only on particular occasions, i.e. holidays. Communicative memory is not constrained by such ritual limitations.

Level	Time	Identity	Memory
inner (neuro-mental)	inner, subjective time	inner self	individual memory
social	social time	social self, person as carrier of social roles	communicative memory
cultural	historical, mythical, cultural time	cultural identity	cultural memory

Figure 1

Table 6: Time and Identity in communicative and cultural memory (in Assmann 2008:109)

At this point, it seems useful to return to our initial decomposition of participation roles, and reflect how these relate to the various knowledge domains just illustrated. We have seen that text recipients will try to assess the coherence of messages by recourse to their previous knowledge which notably consists of different patterns of linguistic, immediate and declarative information. Note that the specific composition of knowledge types can be expected to differ on a personal basis. A child of eight, for instance might easily cope with the technical terminology of joysticks and cartridges used in computer game manuals while it might be completely oblivious to the meanings of lexemes such as “detergent” or “fabric softener” necessary to grasp the gist of manuals for spin driers. Therefore, it is not surprising at all to find text authors adapt their texts to the knowledge they assume is shared with a target audience. To the same extent, ratified readers can also hope to have specific cognitive advantages over non-ratified readers in reading these texts. In other words, ratified readers are “in the know”, so to say and their additional knowledge can come in multiple varieties (e.g. episodic memory, expert knowledge, etc.).

<sup>49</sup> Assmann (1995:128) himself mentions “texts, images, rites, buildings, monuments, cities and landscapes [as] forms of objectivised culture” which coagulate in cultural memories.



	<div> <div>Collective Memory</div> <div>Semantic memory</div> <div>Episodic Memory</div> <div>Communicative Memory</div> <div>Cultural Memory</div> </div>		
Content	recent biographical facts or events	historical events as encountered within the realm of personal biography	petrified, "mythical" history of the absolute past
Form	time & place-bound dynamic, temporary context-dependent	informal, unformed, transient, natural, emergent, rhizomatic	formalized, contingent, symbolic, ritualistic, festive
Media	oral communication (primarily)	oral & written communication	written documents, historical artefacts
Chronological Structure	present - 20 years	80 - 100 years	more than 100 years or past beyond lifetime
Representatives	specific, group of adepts within a discourse community	unspecific, members of a 'community of collective remembrance' ( <i>Erinnerungsgemeinschaft</i> )	specific, individuals capable of defining and engaging in symbolic acts of tradition in any given society

Table.7: The Structure of Memory, based on (Assmann 1992:56)

Some authors purposefully play with the concept of *multiple addressing*. They either leave out or indicate specific information. This effect enables ratified readers to grasp subtle allusions which non-ratified ones would almost certainly miss. While regular ratified users follow weblog entries with interest and have extensive knowledge about its themes, authors and productive contexts, newly ratified or unratified readers may lack the same extent of common ground between weblog author and experienced users.

Sometimes, unratified addressees may even share relevant information with the author which ratified addressees do not. To clarify this issue, an examination of the actual discourse situation is imperative to unravel the exact communication roles of discourse participants. Only then will we be able to discern which piece of knowledge is actually shared by which participants. It is hoped that the empirical

analysis of this thesis will deliver some prototypical examples of such a weblog-specific decomposition of reception roles and disclose users' (lack of) shared knowledge within the aforementioned framework of common ground projected between the blogger and the user(s). Such a divergence of common ground and the confusion of participation roles might amount to users finding some of the entailed content to be unconnected or incoherent (*disturbed coherence*).

As a result, bloggers will thus have to carefully balance their weblog content between exclusive news for "insiders" and inclusive stories fit for a wide range of interested users. While the former orientation is likely to be appreciated by a restricted group of like-minded users, the latter alignment will intrigue a larger readership and enhance its universal appeal within the blogosphere. Weblogs will thus have to choose their place between specification and generalization.

From a productive viewpoint, bloggers will account for their text's content, composition, text design and internal connectivity. Likewise, popular weblogs have to cope with a vast readership whose receptive roles will diverge considerably. To adequately bridge this cognitive divide between "regulars" and "newbies", the hyperlink as cohesive devices makes its entrance. In weblogs, there is no substitute for hyperlinks when it comes to relating the continuous flow of information to specific context-related information. Serial narratives, which stretch along several weblog entries, can be made accessible to weblog "newbies" via hyperlinks which connect previously disclosed story information to a current weblog entry. It has become a regular practise for weblog readers to access serial weblog content in a backward fashion, beginning at the most recent post, and moving backward in time to the very first entry which introduced a certain topic or state of affairs. With some justification, we might thus posit that hyperlinks are "the electronic glue of weblog texts". They trigger connections between entries and comments across the weblog. I shall at a later stage return to these aspects and examine more closely the types of knowledge and participation roles involved in personal weblogs. For now, I shall proceed with the compilation of methodological tools by which I shall analyse the distribution of cohesion in weblogs.

## The Texture: Verbal Cohesion

---

### 4.1 A Framework for Verbal Cohesion in Weblogs

In the following, I will devise a tailor-made framework for the study of verbal cohesion in weblogs. Most categories used in the framework are either enhanced categories based on Halliday & Hasan's influential framework for cohesion published in 1976 as well as more recent work by Tanskanen (2006) and Schubert (2008).

A common technique to establish coherent discourse is the search for interclausal grammatical and lexical continuities in texts. In this vein, we can distinguish four different types of continuities:

- (a) person / agent continuity (personal, demonstrative, comparative reference)
- (b) lexical continuity (lexical cohesion)
- (c) temporal continuity (tense and aspect)
- (d) logical / argumentative continuity (conjunction, rhetorical patterns).

Levels (a) and (b) can be seen to operate on similar grounds. Both levels essentially conjoin discourse elements by ways of lexical identity (a) or lexical similarity (b). On this basis, it seems sensible to converge both levels into one superordinate domain called *topical continuity*. In contrast, levels (d) *logical or argumentative continuity* involves the use of single cohesive devices to elicit the organization or logical connectivity of larger segments of a discourse (cf. figure 19).

While topical continuity is expressed by relations of referential identity and similarity, temporal continuity addresses the formal recurrence of tense or aspect across a stretch of discourse (e.g. walked, shot, went, {PAST TENSE}). Finally, logical continuity relates to the rhetorical relations between clauses often rendered explicit with the help of conjunctions. In this study, we will particularly focus on topical and logical continuities in texts. A crucial reason for this choice is not only the fact that lexical relations "have an important role to play in discourse structure" (Martin 1992:271) but that this type of cohesion has been shown to contribute largely to the creation of coherent discourse. Topical continuities are established by discourse relations between interclausal elements of the discourse. These parts are either

located within or beyond the confines of a clause. Martin (1992) calls clause-internal topical relations (e.g. congruence) *structural ties*. These relations create necessary and relevant parts of the discourse texture. In this study, however, we shall exclusively consider *non-structural* relations, i.e. cohesive relations which exceed the clause boundary (Martin 1992).<sup>50</sup>

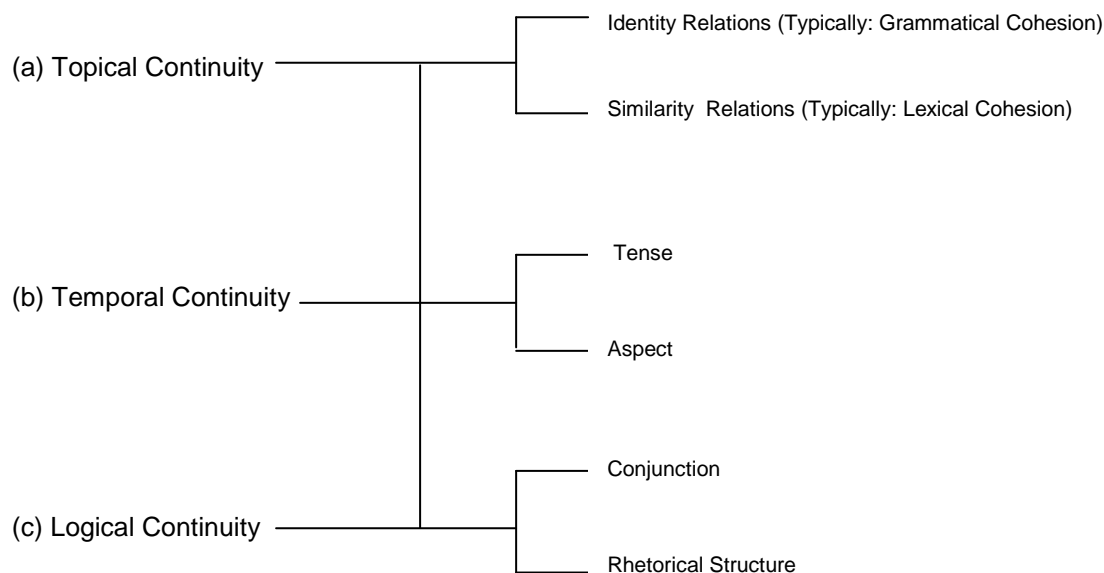


Figure 19: Cohesive Continuity in Discourse

## 4.2 The Scope of Cohesive Relations

There are cohesive studies which focus on structural relations in discourse while others center on non-structural relations. It therefore seems necessary to explain exactly what I consider to be the actual unit of analysis in this study. As the investigation aims to focus on cohesive relations, it seemed sensible to take the cohesive cue as the primary unit of analysis, simply because it triggers the cohesive relation. Thereafter, it was important to define exactly the cohesive range of the cohesive cue. Three methodological possibilities presented themselves:

1. cohesive relations between sentences (*intersentential relations*)
2. cohesive relations between clauses (*interclausal relations*)
3. cohesive relations within clauses (*intraclausal relations*).

<sup>50</sup> The ensuing examples are excerpts from the AWC (Augsburg Weblog Corpus) which has been composed specifically for the purpose of this study. All examples not taken from the corpus will be indicated differently. Anaphoric or cataphoric cohesive cues will appear in bold letters while antecedents or successive referents will be underscored.

Each option refers to a different scope of cohesion. They entail different degrees of complexity. Option number one comprised only relations between sentences but not within or between clauses. This is a very narrow definition of cohesion. Option number three included all kinds of cohesive relations: within the clause, between clauses and between sentences. This seems a very broad definition of cohesive relations. More appropriate than options one and three is option two, which excludes intraclausal relations but accounts for interclausal and intersentential relations. This is the scope which appears to be most useful for the purpose of this study.

I thus decided to exclude intraclausal relations from the analysis since these relations primarily fall into the domain of syntax, i.e. *congruence*, (*lexicographic*) *collocation*, etc. They do not, however, extend clausal boundaries and thus do not actually qualify as discourse relations. Rather, I chose to include both interclausal and intersentential relations in this study since these two form discursive bonds which stretch beyond the clause.

In this context, I would like to point out that previous studies on cohesion have not distinguished sufficiently between interclausal and intersentential relations (cf. Halliday & Hasan 1976, Tanskanen 2006). But it is crucial to differentiate a *sentence* from a *clause*. Both concepts are defined through different criteria. A sentence is usually described as a written unit of discourse which includes some kind of formal closure (i.e. pause, full stop, spacing, etc.). A clause, on the other hand, is not defined through similar formal means but through its particular syntactic function (i.e. the presence or absence of a subject and predicate). By reference to Halliday (1985:192-193), Tanskanen (2006:84f.), for instance, informs us that

the term *clause complex* instead of sentence could be used for the grammatical unit in both spoken and written language, so that sentence could be restricted to refer to the orthographic unit between full stops in written language.

It seems peculiar why we should use the specific term *complex clause* for the basic unit of cohesive analyses, when it would be much more convenient to generally apply the concept of *clause* for that matter. Tanskanen (2006:85) proceeds to maintain that “the sentence rather than the clause” should be selected as the basic unit, arguing that interclausal relations are somewhat “less noticeable” (Tanskanen 2006:85). It remains unclear what Tanskanen implies with “less noticeable” in this context, not least because she ultimately concedes to use interclausal relations herself in her

study. These relations “make the unity of a sentence clearer” (Tanskanen 2006:85). The only reason, therefore, to include interclausal relations in the analysis must be the fact that they are noticeable after all. In addition, Tanskanen neglects to contrast interclausal from intraclausal cohesive relations.

I suggest that cohesive scope is the most promising principle for the differentiation of these two types of cohesive relations. This means that in the analysis, a cohesive cue might thus form a bond with antecedents (or succeeding items) which are either located in the two preceding or following clauses (preceding or following compound or complex sentence).

### 4.3 Grammatical Cohesion

Non-structural relations establish cohesive ties between elements of successive clauses or utterances.<sup>51</sup> Following Halliday & Hasan (1976:6), these relations can be divided into two different branches, i.e. *grammatical* and *lexical cohesion*. I will first introduce and discuss the notion of grammatical cohesion and then proceed to explain its difference with respect to lexical cohesion.

Halliday & Hasan (1976) introduce the concept of grammatical cohesion at the beginning of their study “Cohesion in English”. They argue that grammatical cohesion emerges whenever

the INTERPRETATION of some element in the discourse is dependent on that of another. The one PRESUPPOSES the other in the sense that it cannot be effectively decoded except by recourse to it.  
(Halliday & Hasan 1976:4)

To this end, consider following example serves to illustrate this essential idea of cohesive bonding:

- (1) [...] I waved cheerfully at Claude behind the bar. **He** came out to take our order and I introduced him to Brooke. It took him a while to understand her name.  
(OMW/050708/htqs)

---

<sup>51</sup>Note the underlying semantic relations involved in componential cohesion largely draw on Lyons semantic *sense relations* (1977:270ff).

In (1) the personal pronoun “he” is in initial position of the second clause. In order to identify the reference of the pronouns, we first need to identify another discourse element located in the preceding clause (here: “Claude”). The reader thus needs to establish a cohesive tie between the *anaphor* “he” and its *antecedent* “Claude”. In establishing a cohesive relation between the two discourse items, the reader can thus confirm the topical progression of the noun phrase. Note that the *cohesive tie* (Claude-he) is located within the ongoing discourse. Consequently, the underlying cohesive relation can be termed *endophoric*. Endophoric relations can be subclassified in regard to their directionality. Is the relation prospective, we call it cataphoric. Is the relation retrospective, it is named anaphoric. In example (1) the pro-form “he” points backwards toward the discourse item “Claude”; the underlying relation can therefore be called *anaphoric*. Obviously, anaphoric relations are much more common in most text types than cataphoric ones. Example (2) illustrates such an unusual cataphoric bond between the personal pronoun “him” and its referent “Claude” in a slightly reversed take on the action expressed in (1):

- (2) [...] I waved cheerfully at **him** behind the bar. Claude came out to take our order and I introduced him to Brooke. It took him a while to understand her name.

In (2), it is the personal pronoun “him” whose interpretation depends on the following item “Claude”. We can retrieve the latter succedent from the ensuing clause.

Within grammatical cohesion, there are four cohesive categories which follow this basic working definition of componential cohesion, i.e. *reference*, *substitution*, *ellipsis* and *conjunction*. As such, they all define a semantic relation which underscores the cohesive ties of elements or parts within a given discourse. All of these subcategories will now be introduced consecutively.

#### 4.3.1 Reference

Reference is the first subcategory of grammatical cohesion. It consists of three subtypes, i.e. *personal*, *demonstrative* and *comparative reference*, typically represented by the following verbal resources:

<i>Category</i>	<i>Types</i>	<i>Forms</i>
<b>Personal Reference</b>	Personal Pronouns Possessive Pronouns Possessive Determiners	<i>me, mine, she, her, he, him, his, it, they, their, them</i> <sup>52</sup>
<b>Demonstrative Reference</b>	Demonstrative Pronouns Demonstrative Determiners Adverbs (local / temporal)	<i>this, that, these, those, here, there</i>
<b>Comparative Reference</b>	Pronouns and Adverbs of Identity and Difference	<i>same, equal, such, similar, other, different, adjective comparatives (inflections {-er}, suppletives, e.g. {worse}, periphrastic, e.g. {more important})</i>

Table 8: Personal, Demonstrative and Comparative Reference

Examples (3) and (4) illustrate the cohesive usage of personal reference.

- (3) Anyway, there I was with my mum and my favourite teacher, Miss Richardson  
**She** was, in fact, the only decent teacher we had in the whole school!  
(BMNR/280708/mm)
- (4) Happy 1st birthday to Kathy of the Junk Drawer. Not only is **her weblog** one of the best and funniest around, *she* is a very lovely person, and I am pleased to call *her* my blogging friend.  
(BMNR/300708/hbty)

Similar to example (1), the personal pronoun “she” in (3) refers back to the preceding noun phrase “my favourite teacher [...]”. In (4), the possessive pronoun “her” relates to the preceding personal pronoun “she” (which, in turn, aligns with the initial noun phrase “Kathy of the Junk Drawer”). Therefore, pronouns refer to preceding or succeeding discourse items by triggering identity relations between the resulting cohesive ties.

More intricate than the default cases of personal reference is the cohesive force of possessive and demonstrative determiners. In (4) the possessive determiner “her” in

<sup>52</sup> Note that the interpersonal pronouns, *I, you* and *we* were analysed as part of lexical cohesion (total or partial recurrence). They are thus not included in this list.



the second clause modifies the noun phrase “her weblog”. Determiners are always modifying parts of underlying noun phrases, and it is legitimate to ask whether a cohesive tie is triggered by a determiner (personal reference) or the head of its noun phrase (lexical cohesion). In our case, the possessive determiner evokes the possessive tie to the preceding noun phrase “Kathy of the Junk Drawer”, not the noun “weblog”. The cohesive force primarily rests on the determiner alone. On these grounds, the cohesive use of possessive determiners can be readily subsumed under the category of personal reference. Let us now take a look at the cohesive force of demonstrative reference.

- (5) The photo below shows the tidiest room in the house - only because **this** is where we had to stay all day!  
(BMNR/100708/oln)

- (6) The box smirked at us again. The only way we were going to get **this box** was by sliding the thing off the shelf and onto the trolley.  
(BMNR/130708/wsofhodw)

In (6) the demonstrative determiner “this” correlates with the adjacent noun “box”. This time it is the noun which principally refers to the noun phrase “the box” located in initial position of the preceding clause, not the determiner (as was notably the case in (4)).

Halliday & Hasan (1976:58ff.) count demonstrative determiners as regular forms of the reference category, arguing that the use of demonstrative determiners indicate specific proximity relations between discourse items which are near (*this, these*) or distanced (*that, those*). It seems obvious that proximity indications of this sort are only viable if the resolution of a cohesive tie problematic to the hearer/reader and the cohesive tie cannot be resolved (possibly because there is more than one possibly antecedent and the anaphoric relation is ambiguous). However, in the case of example (4), such a situation does not present itself. Rather, the cohesive relation is resolved easily on a mere lexical basis, connecting the noun “box” to its repetitive item in the following discourse. I suggest that the cohesive relation in (4) is part of lexical cohesion (see 5.3.1.2) and not of demonstrative reference.

Demonstrative determiners can trigger reference relations indicating the proximity of a related discourse item. Note that from a cohesive point of view, such indications are only useful if they help to clarify an ambiguous anaphora. In such cases, demonstrative determiner is a formal realization of the reference category.

A regular cohesive use of the demonstrative pronoun “this” is illustrated in example (7). In contrast, see example (8), which again exhibits a demonstrative determiner rather than a pronoun setting up the anaphoric tie (this show – the segment). Realize that in (8) the anaphoric resolution of “this show” is more problematic than the one in (5). Firstly, the head of the noun phrase “show” is not formally repeated in the preceding discourse and secondly there are multiple noun phrases in the preceding clause, which possibly qualify as antecedents, e.g. *a segment*, *Guinness diet*, *Bob’s challenge*.

- (7) Recently I read something amazing in one of the 9,458 political Weblogs that I follow. It was written by a guy, and so, you won't be surprised by the sports reference. (Men will never tire of **this**. I don't know why.)

(OMW/290708/twobmtqos)

- (8) Our pal Fiachna played a significant role in a segment a few months ago called the Guinness Diet. Bob's challenge for **this show** was to survive for one solid week on a Guinness-only diet.

(OMW/030708/naesvp)

We can argue that the hyponymy relation between the lexical pair “show” and “segment” primarily secures the cohesive force of the tie. In contrast, the demonstrative determiner supports the cohesive relation but does not constitute the relation. Nevertheless, the demonstrative determiner fulfils the important cohesive function of indicating the local proximity of the cohesive tie. In (8), it does not only boost the relation between “show” and “a segment” but likewise sets up a second cohesive relation between “show” and “the Guinness Diet” which is closer to the anaphor. Therefore, in our example, at least, it maintains a secondary cohesive relation on its own.

Another difficult case is the cohesive status of the definite article as illustrated in example (9):

- (9) [AE] told her that he was praying to God to take away all the mean mommies and daddies. Oh, **the poor child**.  
(NSM/070708/aws)<sup>53</sup>

One could argue that the definite article here serves the purpose of designating the interclausal relation between the noun phrase “poor child” (which it modifies) and a noun phrase in the preceding clause (“AE”). The article is then only indicative of some antecedent in the previous stretch of discourse whose character remains unclear. In fact, multiple noun phrases are possible referents: AE, her, GOD. Referential clarification is again achieved primarily by means of a lexical relation between “child” and “AE”. It is the lexical meaning of “child” which clearly enables us to select the most plausible antecedent “AE”. Similar to the cohesive force of demonstrative determiners, definite articles only underscore an existing lexical tie. They may be reinforcing the lexical tie by grammatical means but do not specify the nature of the antecedent such as person and gender (possessive determiners) or demonstrative determiners (proximity). This marks the cohesive surplus of determiners. Not surprisingly, Halliday & Hasan (1976:71) distinguish definite articles from determiners which they hold to be “semantically selective”. In contrast, they describe definite articles as “[having] no content” (Halliday & Hasan 1976:71). The anaphora “the poor child” can thus be categorized as an example of lexical cohesion (paraphrase). Unlike Halliday & Hasan (1976), definite articles will thus not be deemed formal realizations of the reference category in the course of this study.

Let us now return to example (10). Interestingly, the anaphoric resolution of the noun phrase “the ending” seems more demanding than in the previous examples. Yet, the noun phrases “the ending” and “the matrix” interlace on a cohesive plane. Surely therefore the definite article must signal a different sort of cohesive relations than classic lexical ties. In this case, the blogger of this passage probably expected her users to possess and “fill in” missing background information necessary to form the cohesive tie between the two noun phrases. The “missing link” can perhaps be described along the lines of social knowledge (that the blogger regularly talks about personal film experiences in movie theatres), procedural knowledge (about film plots

---

<sup>53</sup> The letters AE in the example represent the initials of a close relative of the author, whose name was introduced in previous postings.

and their structure, development), (pop) cultural knowledge (about recent or current films and their titles). If the user has access to all (or some) of these pieces of information prior to reading the passage, he can apply them to cohesively *bridge* the two noun phrases in question, i.e. “the ending” and “the Matrix” (cf. Matsui 2000). In other words, the author presupposes information because he thinks the readers of his text will be capable of “bridging” its parts. Readers’ implicit knowledge thus joins up an otherwise ambiguous anaphora (or cataphora) to the contextually most suitable element.

This is no longer an example of cohesion but already of coherence because readers need to rely on their socio-cultural knowledge to decode the implicit discursive relation.<sup>54</sup> The example shows that there is a thin line between the cohesive indication of thematically-related discourse parts and the pragmatic reconciliation of knowledge and ongoing discourse.

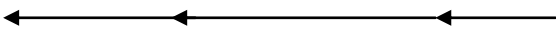
Generally speaking, one semantic category, namely reference, could be introduced with the help of different examples. There are many different formal realizations of it in actual discourse, e.g. pronouns or determiners, and we should be careful not to confuse the functional (cohesive categories) with the formal level of analysis (cohesive forms).

The conceptual difference between functional category and formal realization has some additional implications: imagine a noun phrase is introduced in a larger unit of text, starting with the words “the President of the United States”. Assume that this phrase is recurrently addressed throughout the ensuing discourse by various formal reifications, e.g. personal, possessive, demonstrative pronouns. In this case, it seems unreasonable to expect readers to remember at all times the exact wording, in which the initial noun phrase originally presented itself at the beginning of the text. More important than recalling the exact wording of a referent is the construction of a mental representation of it to which all potential anaphoras can be referred. There is no need for actually remembering the form of a discourse item as long as it occupies a specific referential structure in our mental representation of the discourse. In example (10), for instance, we can only grasp the successive conceptual changes of the noun phrase “chicken” if we do not recognize identity on a formal but on a more abstract, relational and cognitive plane. What we need are not necessarily formal equivalences

---

<sup>54</sup> Matsui (1993, 2000) calls such cognitive operations *bridging*.

but semantic and pragmatic analogies. Therefore, we can conclude that cohesion is a semantic as well as pragmatic concept:

- 
- (10) Kill an active, plump chicken. Prepare **it** for the oven, cut **it** into **four pieces** and roast **it** with thyme for 1 hour.  
(Brown/Yule 1983: 202)

Although the initial referent “chicken” obviously alternates, readers are able to keep track of the referential change, adapting their concepts as the discourse unfolds. Again, I wish to stress the fact that cohesive relations are therefore not to be found *in* the discourse, as it were. Rather, in hearing or reading we ascribe cohesive relations to the discourse. Cohesion depends upon our own ability to map discourse elements or units onto each other through a referential scheme of mental representation (cf. also Brown & Yule (1983:200):

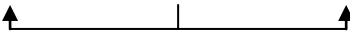
It needs to be stressed that pro-forms, whether functioning anaphorically or cataphorically, are tokens of meaning which only make sense when their relationship with what has preceded or what follows, is both identified and interpreted.  
(Widdowson 2008:47)

There is yet a third category proposed by Halliday & Hasan (1976), namely comparative reference. Discourse-based comparative relations are inherently referential in character:

Likeness is a referential property. A thing cannot just be ‘like’; it must be ‘like something’. Hence comparison is a form of reference.  
(Halliday & Hasan 1976:79)

Following Schubert (2008), comparative reference can be expressed by virtue of identity markers (*same, equal*), similarity markers (*such, similar*), disparity markers (*other, different*), comparative adjectives (e.g. *better, worse, smaller*) or modifying adverbs (*so, as*). With a view to its formal realization, comparative reference differs to some extent from personal and demonstrative reference. As seen above (personal and demonstrative) reference is expressed by cohesive ties of which one item acts as cohesive cue (anaphora or cataphora) pointing towards the other (preceding or succeeding) element. In contrast, comparative reference consists of one cohesive

cue (comparator) which is not part of the cohesive tie it indicates. Rather the comparator connects one preceding and another following discourse element or segment. Comparative reference is thus realized by a tripartite complex of elements as expressed by arrows in (11) with *such* being the comparator connecting the two different noun phrases (underscored).

- (11) My Tammy Fay Baker eye makeup smudged with yesterday's pitiful black-and-blue attempt at creating the illusion of come-hither eyes. Sore, I was, in unmentionable places. **Such** as my calves.
- 
- (OMW/290708/twobmtqos)

Put briefly, in comparative reference, the cohesive cue is no longer part of an independent cohesive tie but merely indicates the implicit connection between two further discourse parts or segments (see figure 22). We can thus distinguish between *internal* and *external* cues. An internal cue refers to cohesive relations of which one element acts as cohesive cue. I call this relation internal because the underlying cohesive relation is made explicit *within* the cohesive tie. Cohesive cues are external when they do not comprise a cohesive device but need a third independent discourse component for their indication.



Figure 19: Internal and External Cohesive Ties

#### 4.3.2 Substitution and Ellipsis

As part of grammatical cohesion, Halliday & Hasan (1976) discuss *substitution* and *ellipsis* on par with personal, demonstrative and comparative reference. Basically, substitution refers to cohesive relations which replace a preceding discourse segment with a specific lexical item.

- (12) Here are some pictures of our relaxing days. In the 2nd **one** you can kind of make out someone on the zip line over the pool.  
(FB/070708/crhp3)

In example (12) the lexical item *one* stands in for the preceding nominal phrase “some pictures” constituting *nominal substitution*. Likewise, the lexeme *same* is can substitute nominal phrases. In (13), for instance, the lexeme “same” replaces the preceding nominalization “punishing EVE”.

- (13) Guilty as sin. As original sin. If it's ok for God to punish Eve, then it's ok for men to do the **same**, to infinity, and beyond.  
(OMW/290708/twobmtqos)

In contrast, *verbal substitution* makes use of the verb *to do* to replace verbal processes (cf. examples (14), (15)).

- (14) Who knew I had it in me to be so prolifically snide. Well, I **did**, I guess  
(AA/190708/oes)
- (15) Erica's picture (alone because Mr. X will not allow pictures on the weblog...) would look a lot less awkward if he **did** - hint, hint).  
(FB/280708/50<sup>th</sup>a)

Finally, *clausal substitution* can be achieved with the help of the lexical items *so* and *not*, which respectively represent a preceding clause rather than a nominal group. Example (16) and (17) illustrate how clausal substitutions work. Note that in example (18) the cohesive marker *so* functions as a nominal not as a clausal replacement. Actual replacements of some substitution markers thus need to be assessed in their context of use.

- (16) I am aware that I could *mow the lawn myself* and actually was planning to do **so**, but when I mentioned this to my mother she apparently deemed it unacceptable.  
(NSM/300708/tpbtybatlos)

- (17) Maybe these accomplishments will mean she'll start sleeping better sometime soon. Or judging from last night, maybe **not**.  
(NSM/070708/aws)
- (18) The building was built between 1903 and 1905 as a home for the New York Times, and it remained **so** until 1961.  
(MM/010708/tsa)

The cohesive category of *ellipsis* relates to a “substitution by zero” of related nominal, verbal or clausal discourse segments (Halliday & Hasan 1976:142). Ellipsis thus constitutes a structural omission of certain discourse parts immediately retrievable from the preceding discourse. Consider (19) and (20) for two instances of nominal ellipsis:

- (19) I'll put the rest in this weblog, and [I] will try not to make it too long.  
(FB/070708/crhp3)
- (20) In fact I got two rings. I wasn't in to diamonds, so I got two gold rings with garnets. My favourite stones. Both [**rings**] were very old and unusual designs.  
(BMNR/200708/atdml)

Examples (21) and (22) present possible verbal cases of ellipsis. In (21) a potential ellipsis specifies the implicit subject and operator verb (copula “be”) of the embedded clause.

- (21) [**I am**] Thinking of you!  
(OMW/130708/som/Co2)
- (22) I've decided that I'm going to take one picture a day for one whole year. I'm going to try to find a theme to do each month, but I need one to start off with.  
[**Do you have**] Any suggestions?  
(FB/050708/pedfay)



It is difficult to determine whether the two previous ellipses in (21) and (22) are actually not cohesive examples at all because we do not have recourse to their actual surrounding discourse. Provided that there is no formal resolution of the elliptical content of the clauses within the discourse, we can only suspect that the ellipses are resolved by recourse to individual (grammatical) knowledge. As such, the two previous examples should be allocated to the category of coherence and not cohesion, i.e. bonds between the discourse and individual memory. Example (23), on the other hand, illustrates a regular case of a clausal ellipsis including the full exclusion of clausal segments which are, however, fully recovered from the immediate co-text of the remark:<sup>55</sup>

(23) And I don't often find sea glass. Plastic bottles, yes **[I do find often]**. Sea glass, no **[I do not find often]**.  
(AA/050708/stp/Co13)

As we have seen before, cohesion sometimes affords a third discourse element which bridges two implicit discourse elements of segments (e.g. comparative reference). Now, we can distinguish three different cohesive categories (reference, substitution and ellipsis) and two different formal reifications (explicit and external cues).


Reference	Substitution	Ellipsis	
personal	nominal	nominal	} <i>internal cues</i>
demonstrative	verbal	verbal	
	clausal	clausal	
comparative			} <i>external cues</i>

Figure 20: Preliminary Set of Cohesive Categories (Grammatical Cohesion)

<sup>55</sup> In Givón (1990:399ff.) we can learn that this difference also applies to the case of referential nominals which are identified “by a speech community because it represents widely shared knowledge [mnemonic relations], by its immediate deictic availability in the speech situation [exophoric relations] or by its being mentioned in prior discourse [endophoric relations].”

### 4.3.3 Conjunction

The fourth cohesive category of grammatical cohesion can now be introduced, i.e. *conjunction*. With this cohesive marker, interlocutors indicate semantic relationships not between lexems but clauses. The category of conjunction is realized either by adverbials or conjunctions.<sup>56</sup> According to Halliday & Hasan (1976) four general types or kinds of conjunctive relations exist, i.e. *additive*, *adversative*, *causal*, *temporal* (cf. Schubert 2008: 42-22). Conjunctive relations include one cohesive device connecting two interrelated discourse segments (a preceding and a following clause) to establish one formally external cue:

- (24) OK, I was honestly sort of coveting this award when I saw it going around. I know coveting is a deadly sin but I was doing it anyway. **Then** out of nowhere RED comes to the rescue. Thank you,  
(WIT/150708/maa)
- 

The following three examples (25), (26) and (27) give ample evidence of conjunctive clause relations in my weblog corpus (e.g. “and” – additive, “but” – adversative, “so” – causal, “then” and “finally” – temporal).

- (25) [...] **and** William told That Guy that he wouldn't be in the shoot that evening, **but** would be in the next evening's shoot with all the French food critics. **So**, he trundled on home and I became Fiachna's date.  
(OMW/030708/naesvp)
- (26) **then** I pushed my right hip out and then the left hip out and shook it all about! **Finally**, I lay my head forward and tried to rest my head on the headrest of the person in front of me.  
(MM/260708/itfft)
- (27) I also said I didn't eat cream and these are full of fresh cream.  
(BMNR/020708/ilodi)

There is an ongoing dispute in discourse analysis whether cohesive conjunction subsume “paratactic and hypotactic [...] linkers within sentences” (Martin 2001:36). In

<sup>56</sup> There is, as you can see, a difference between the cohesive category of *conjunction* and its possible formal realizations as (word class) *conjunctions*.

this study, I take the position that cohesion is primarily an interclausal phenomenon. Hence, a conjunctive relation located between a superordinate and a subordinate clause within an underlying complex sentence will be considered cohesive. Similarly, Gutwinski (1976) proposed to count all conjunctive indicators connecting hypotactic and/or paratactic clauses as cohesive devices.<sup>57</sup> In summation, the cohesive categories of reference, substitution, ellipsis and conjunction can be referred to as *grammatical cohesion* because they “involve small, closed classes of items or gaps” (Martin 2001:36). The second important part of componential cohesion is called lexical cohesion. It refers to the formal and conceptual reiteration of lexical items and shall be presented in the following.

<u>Reference</u>	<u>Substitution</u>	<u>Ellipsis</u>	<u>Conjunction</u>	
personal demonstrative	nominal verbal clausal	nominal verbal clausal		} <i>internal cues</i>
comparative			additive adversative temporal clausal	
				} <i>external cues</i>

Figure 21: Full Set of Categories of Grammatical Cohesion

#### 4.4 Lexical Cohesion

With lexical cohesion we leave the descriptive trail laid down by our initial working definition of Componential cohesion. Let us recapitulate once more its most basic claims:

the INTERPRETATION of some element in the discourse is dependent on that of another. The one PRESUPPOSES the other in the sense that it cannot be effectively decoded except by recourse to it.  
(Halliday & Hasan 1976:4)

Accordingly, any discursive item must – in order to be cohesive – be interpretable only by recourse to another discourse segment or element present in its surrounding

<sup>57</sup> The analysis of theme-rheme progression will not form part of this study simply because they do not relate either to the two-fold structure of explicit cohesive ties, nor to the complex configuration of external cues.

discourse. As we will see, this limited definition does not strictly apply to examples of lexical cohesion:

(28a) On Monday, my new pal Brooke called me from Charles de Gaulle airport and asked me if I could log in to her email and check her mother's itinerary, because *Mom* didn't get off the plane. Of course! What are friends for?  
(OMW/050708/htqs)

(28b) On Monday, my new pal Brooke called me from Charles de Gaulle airport and asked me if I could log in to *her* email and check her mother's itinerary, because Mom didn't get off the plane. Of course! What are friends for?  
(OMW/050708/htqs)

If you compare the personal reference in example (28b) with the lexical cohesion of the synonyms “Mum” and “mother” in (28a), it becomes apparent that in order to grasp the conceptual meaning of the possessive determiner “her”, you have to retrace the preceding noun phrase “Brooke”. Thus, the cohesive marker “her” presupposes “Brooke” in the way disclosed in the above definition. This is obviously different in (29a), where the two noun phrases plainly stand in some recognizable semantic relation to each other but do not need to be correlated for an effective interpretation of the anaphora. In (29a) a conceptual interpretation for each element of the cohesive tie is possible without prior interclausal connection. In fact, in lexical cohesion, cohesive connectivity succeeds individual interpretation. This distinction between grammatical and lexical forms of cohesion was already discovered by Hoey (2005:71) who states:

Lexical items in lexical repetition or paraphrase links do not depend on each other for their meaning (at least not in the definitional sense), though they share a dependence of co-text for realization of their meaning potential. Textual (grammatical) items, on the other hand, entirely depend on other items for their interpretation and have no definitional meaning in themselves

This new set of lexical cohesion consists of six elemental categories of cohesion based on different degrees of conceptual similarity.<sup>58</sup> These are

- Repetition (Total or Partial Recurrence)
- Equivalence (Synonymy, Paraphrase, Parallelism)

---

<sup>58</sup> For a comprehensive list of lexical categories of cohesion in previous research on cohesion, see Tanskanen (2006).

- Superordinate (Hyponymy, Hyperonymy, Holonymy, Meronymy)
- Antonymy (Complementary, Contrary, Converse Opposition)
- Co-Hyponymy
- Collocation (Activity-Related and Elaborative Collocation).

#### 4.4.1 Repetition (Total and Partial Recurrence)

Repetition of lexical items or segments occurs either by total formal recurrence (total recurrence) or by partial morphological reiteration of the word stem or base, allowing for various formal alternations of a cohesive discourse item (partial recurrence):

(29) And you [...] will be happy to know that I couldn't have another cigarette if somebody paid me to smoke. There's some link in my brain between *smoking* and puking now, and so, I can't even look at *cigarettes*.

(OMW/050708/htqs)

(30) Yesterday I left a comment on JD's [...] post about chocolate. I said I hadn't eaten *chocolate* in years....

(BMNR/020708/ilodi)

(31) Going to this beach always does me good. It *does* all of us *good*.

(AA/090708/avdc)

It is important to point out that the cohesive relation involved in the category of repetition is structurally, not semantically induced. The analogueical relations gained from its discourse relation are motivated by perception rather than conception. Example (29) provides two cohesive ties between the items *cigarettes* and *cigarette* and *smoking* and *smoke* based on partial recurrence.<sup>59</sup> (30) and (31) present the case of full recurrence with the cohesive ties (*chocolate* – *chocolate*, *does* – *does*, *good* – *good*).

#### 4.4.2 Equivalence (Synonymy, Syntactical Parallelism, Paraphrase)

<sup>59</sup> Recurrence between lexical discourse items can be further divided up into categories explored by Halliday & Hasan (1976:283), i.e. *exclusive*, *identical*, *inclusive* or *unrelated reference*. These differentiations reflect the fact that recurrent discourse elements may or may not refer to the same person, object, concept or entity in the world while they both have the same extension.

Equivalence involves two lexical items which share semantic or syntactic similarity. The cohesive ties involved in equivalent discourse relations can either be represented by partial synonyms (e.g. horse – stallion), similar syntactic constructions (parallelism) or by complex descriptions (paraphrase). Example (32) illustrates an instance of synonymy between the anaphora “mom” and its antecedent “mother”. Here, similarity between the items is thus strictly conceptual, based on semantic grounds.

- (32) On Monday, my new pal Brooke called me from Charles de Gaulle airport and asked me if I could log in to her email and check her mother's itinerary, because *Mom* didn't get off the plane. Of course! What are friends for?  
(OMW/050708/htqs)

In some ways, example (31) can equally be regarded as a form of equivalence. However, in contrast to (32), the similarity between the two clauses is not one of semantic proximity but of syntactic parallelism. Hence it does not imply an equivalence of meaning but rather an equivalence of form. A *syntactic parallelism* relates to similar structural clause patterns (“X does Y good”) which may exhibit different lexical fillings as can be seen in (32):

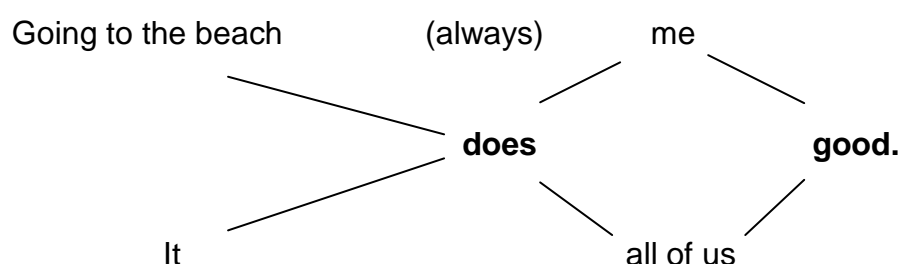


Figure 22: Syntactic Parallelism Across Clausal Boundaries

Syntactic parallelism is an important part of discourse cohesion if its integral syntactic similarity exceeds the clause boundary. Instance of intraclausal syntactic parallelism will be excluded from this study for reasons already mentioned.

Halliday & Hasan (1976) exclude intraclausal discourse relations from their study of interclausal cohesive relations. Following their example, equivalence relations

which extend beyond the clausal realm shall be considered cohesive (see (33), others will be disregarded in this study.

(33) Fiachna called, after spending most of the day (2pm - 7:30pm) at the local hospital, helping out one of the little Irish girls - *a friend of his daughter*.  
(OMW/060708/shs)

(34) The owners, Claude and Zaina, have not updated its classic 19th century bistro style interior,  
(OMW/050708/htqs)

(35) Apparently Earth has a natural soundtrack: *a high-pitched series of chirps and whistles, that could, potentially, be heard by aliens* if they knew how to listen.  
(AA/020708/ens)

Examples (33) and (34) both include intraclausal types of the paraphrase. In contrast, example (35) presents a classic transphrastic example of the phenomenon, the two integral clauses being connected by a colon. Note that in this case, the anaphoric noun phrase “a high-pitched series of chirps and whistles [...]” exceeds in complexity the preceding noun phrase “a natural soundtrack” to which it builds up a cohesive relation. Such relations, in which the secondary, more complex discourse segments elaborates on a preceding noun phrase, are called *expansions*. There is a second subcategory of the paraphrase which we can call *condensation*, following Wolf (1981: 207).<sup>60</sup> It applies to the event in which the first cohesive noun phrase is more complex than the second one, as shown in the following example Schubert presents in this context:

(36) The floorboards creaked, the plumbing whined and throbbed, doorhinges squeaked and windows rattled in their frames. *The noise* was deafening.  
(David Lodge 1981)

---

<sup>60</sup> These latter relations have elsewhere been classified as *shell nouns* (Schmid 1999) or *labels* (Gill 1994) and qualified to have further evaluative and pragmatic meanings beside its cohesive use.

The example shows how the various noises evoked by an assemblage of different verb phrases (creaked, whined, throbbed, rattled) are condensed by the anaphora *the noise* located in initial theme position of the successive clause.

#### 4.4.3 Superordinate (Hyperonymy, Hyponymy, Holonymy, Meronymy)

According to Lyons (1976:291), a *superordinate* is defined by its “relation which holds between a and a more general, or superordinate, lexeme and a more specific, or subordinate, lexeme”. In example (37), for instance, the anaphora “that doctor” refers back to the preceding hyponym *paediatrician*, thereby triggering a superordinate relation between the cohesive tie *doctor-paediatrician*.

- (37) I worked for several different pediatricians [...] I worked for that doctor, a male last. He was also really good to me and to my family.  
(ShSe/220708/cz)

Superordinate nouns which have a general or vague reference, such as thing, place, people, person, object, matter, etc., are sometimes called *general word*. Their versatility makes them very frequent in natural discourse:

- (38) In addition, you have to be prepared to slog through the crowds at the corner of Boulevard Rochechouart and Rue de Clignancourt. I've been through **this place** so many times, that I recognize the Middle Eastern "spotters" that control that 2-cornered territory.  
(OMW/160708/fgfg)

*Hyponymy* figures as the reversed picture of superordination. It entails the relationship between a more specific lexeme towards a more general or superordinate one in the discourse. For instance, example (39) features the noun phrase *the best catcher*, which relates to the preceding noun phrase “the youngest player” in an hyponymic fashion. The underlying semantic bond between the cohesive pair *catcher* – *player* is one of subordination. Moreover, a special case of hyponymy is *meronymy*, which constitutes a part-whole-relation between a specific cohesive pair, e.g. *door* - *house*, *arm* - *body*, *motor* - *car*, *emergency room* - *hospital* (cf. 40a). Note that if we were to change anaphora and antecedent so that “company”



would refer anaphorically to *the department*, the relation would point from the more general item to its part (cf. 40b). This latter relationship is not called meronymy but holonymy.

(39) She is their catcher and probably the youngest player on the team, but she is a joy to watch. She is probably **the best catcher** I have ever seen.  
(ShSe/140708/ppaf)

(40a) Yeah, the company is still the way it was, maybe worse, but **the department** I work in and the people I work for and with are the best.  
(ShSe/220708/cz)

(40b) Yeah, the department is still the way it was, maybe worse, but **the company** I work in and the people I work for and with are the best.  
(ShSe/220708/cz)

#### 4.4.4 Co-Hyponymy

Superordination and hyponymy has seen a superordinate lexeme include semantically the extension of a subordinate one. In contrast, co-hyponyms are lexical items, which are both hyponyms to a superordinate lexeme or concept. They have different extensions but share the same hyperonym, e.g. *fork-spoon* (cutlery), *boy-girl* (children), *bear – tiger* (animal).

#### 4.4.5 Antonymy (Contrary, Complementary, Converse and Directional Antonymy)

The cohesive relation of antonymy signals to the reader a difference in meaning between two opposing lexical items. This antonymic relation forms a cohesive tie and can surface in four different constellations:

- **Complementary Antonymy**

This category defines the binary contrast between two given lexical items in the discourse which are not gradable, e.g. *dead-alive*, *inside-outside*, etc.

- **Contrary Antonymy**

Gradable lexemes which stand in some gradable contrastive relation to each other are called contrary antonyms. Hence, contrary antonymy allows for

comparative relations of the following kind: “A is smaller than B” or “B is larger than A” (*small-big*). The opposition between the two adjectives is not complementary.

- **Converse Antonymy**

This contrastive relationship is characterized by being “the result of a change of perspective” (cf. Schubert 2008:49). Converse antonymy is usually expressed through procedural verbs as well as nouns which express reciprocal social roles, e.g. *buy-sell*, *come-go*, *lend-borrow*, *husband-wife*, *teacher-pupil*.

- **Directional Antonymy**

This last sub-type of opposites refers to lexical pairs which imply “a motion in one of two opposed directions with respect to a given place” (Lyons 1977:281), e.g. *up-down*, *north-south*, *forwards-backwards*, etc.

#### 4.4.6 Collocation

The last (and possibly most controversial) cohesive category of lexical cohesion is *collocation*. The term was originally coined by John Rupert Firth (1957) addressing the frequent co-occurrence of adjacent lexical elements in discourse, e.g. *blonde hair*, *handsome boy*, *a flock of sheep*. Halliday & Hasan (1976:284) build on Firth’s classic intraclausal concept of collocation and extend it to the interclausal realm. They thus establish collocation as a cohesive notion. Their description of cohesive collocation is somewhat vague, referring to “the association of adjacent lexical items that regularly co-occur”. Even Halliday & Hasan (1976:284) admitted to the fact that their broad definition of the category turn into into the “most problematic part of lexical cohesion”. Nonetheless, I claim that collocation plays a vital part in the construal of cohesion in most text genres. On a methodological plane, the definition proposed by Halliday & Hasan (1976) needs to be refined.

Schubert (2008:52) proposes a useful classification of collocation into two different subtypes. He draws on Lipka’s semantic distinction between *lexical fields* and *lexical sets* (Lipka 2002). Lexical fields are lexical chains which share the same word class. In addition, they possess at least one mutual semantic feature. In contrast, lexical sets are “based on association and intuition or on objectively verifiable relationships captured by encyclopedic knowledge” (Lipka 2002:173). In

contrast to lexical fields, lexical sets must not be part of the same word class.<sup>61</sup> Example (41) presents range of lexical elements subsumed to the lexical set of [SMOKING], e.g. *cigarettes*, *gab*, *smoking*, *burn*, *a whole pack (of cigarettes)* and *carton (of cigarettes)*:

(41) Then, we pulled out the **cigarettes** and had a good **gab**. Yes, I started **smoking**, after not **smoking** since 1948, at least. Yes, That Guy blames Brooke. Yes, she did let me **bum** one from her the first time we met for lunch. And yes, she gave me **a whole pack** from **the carton** she brought from California. But in truth, I blame That Guy. Why not? He's not here to defend himself.

(OMW/050708/htqs)

Based on a classification introduced by Martin (1992) and in line with Tanskanen (2006:62), we can further distinguish two relational subtypes, i.e. *activity-related collocation* and *elaborative collocation*. Activity-related collocation occurs when “actions, people, places, things and qualities configure as activities” (Martin 1992). In (41) the cohesive tie (cigarettes-smoking) can be classified accordingly. The frequent co-occurrence of both items is activity-related in the sense that cigarettes are usually smoked as much as food is eaten. The second category of elaborative collocation (Tanskanen 2006:62) is more subjective and basically connects with Lipka’s concept of lexical sets. From Lipka’s definition of lexical sets (or elaborative collocation), we learn that the detection of collocation pairs depends heavily on the knowledge which interlocutors bring to the interpretation of discourse. This cognitive footing aligns collocations (perhaps more than other cohesive relations) with the cognitive level of text production and comprehension. It is, indeed, plausible that webs of collocation relations reflect and induce underlying cognitive frames (Fillmore 1987, Ungerer und Schmid 1996). According to Minsky (1988) a frame is a basic cognitive structure of knowledge which can be described as follows:

[w]hen one encounters a new situation (or makes a substantial change in one's view of the present problem) one selects from memory a structure called a *Frame*. This is a remembered framework to be adapted to fit reality by changing details as necessary.

(Minsky 1988:112)

---

<sup>61</sup> „By analyzing or describing a word is to be understood its analysis in terms of the sense-relations which it contracts with other words; and each such sense-relation can be explicated by means of what Carnap called meaning-postulates” (Lyons 1977:204), e.g. synonymy, repetition, hyponymy, opposition, etc.

In the case of (41), we can posit that the mention of collocational noun and verb phrase (e.g. cigarettes) triggers a specific cognitive frame in interlocutors' minds, i.e. the frame of "SMOKING". Collocations might thus be classified according to the frames they induce over larger stretches of discourse. It is precisely this fact, which renders collocation "a very subjective relation" (Tanskanen 2006). At the same time, I argue that their close relation to frames is a cohesive surplus of collocation. It is precisely this subjective tinge of collocation, which makes it an extremely valuable asset to the study of the dynamic construal of cohesive texture. It performs bridging function lodged between the productive creation of collocation webs in texts and their cognitive frame-based interpretation in discourse. As such, the category thus becomes an important facet of my investigation of lexical cohesion. Figure 23 summarizes the range of lexical cohesive relations which shall be applied in the empirical analysis conducted presently.

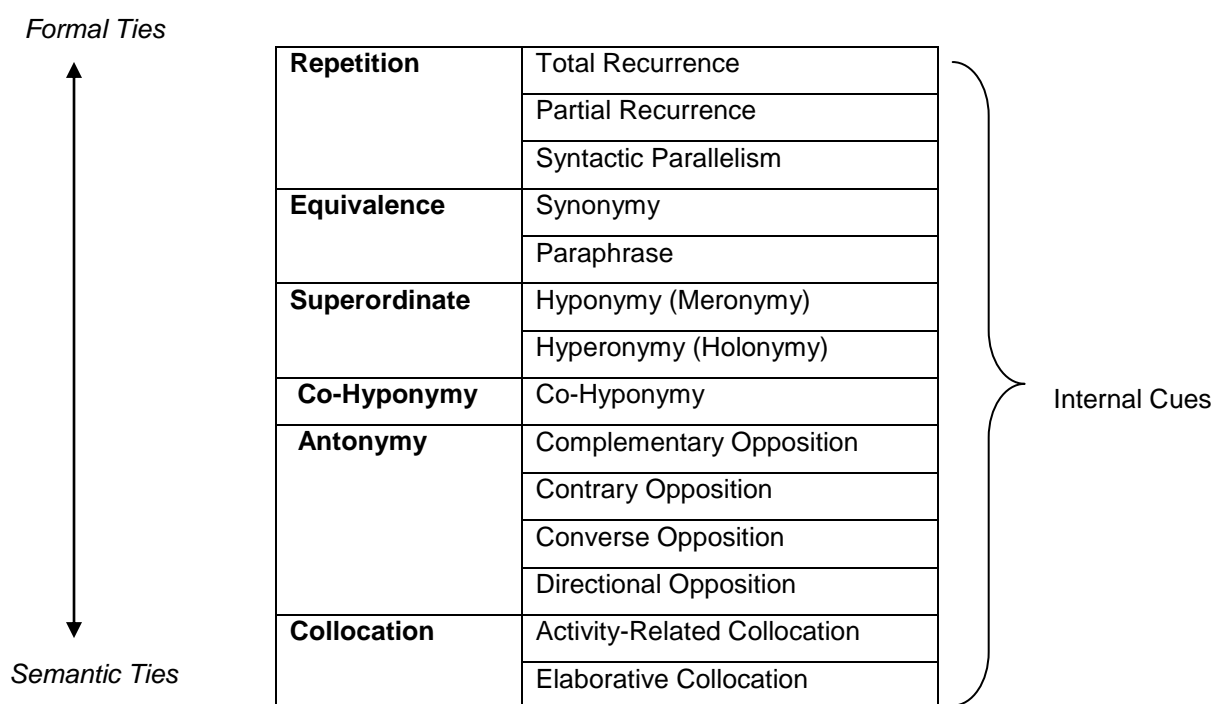


Figure 23: Lexical Cohesion in Verbal Discourse

We can see that while the category of repetition entails at least some degree of formal resemblance between its cohesive ties, all other categories are only based on semantic analogies. As Halliday & Hasan (1976) suggest, lexical cohesion is thus mainly (although not entirely) a semantic concept. Before this framework of categories and relations will be applied to empirical corpus data, it is necessary to first outline how the weblog corpus for this purpose was compiled and prepared. The

next chapter will therefore deal with the methodological process of corpus compilation and the ensuing stages which were involved in its automatic analysis.

## The Corpus

---

The previous two chapters were devoted to the creation of a framework suitable for the description of cohesive relations in weblogs. Chapter seven and eight will deal with the application of the framework to a self-compiled corpus of personal weblogs (AWC).<sup>62</sup> The corpus was assembled in a number of consecutive methodological stages which shall be introduced in this chapter. Chapter nine will reconnect with the topic of discourse, cognition and participation roles introduced in chapter four, applying the theoretical concepts introduced previously to a range of authentic weblog examples. However, let us now first take a look at the way the weblog corpus was compiled, segmented and prepared for the automatic and manual analysis of cohesion.

### 5.1 The Augsburg Weblog Corpus (AWC)

The weblog corpus compiled for this study comprises ten generically compatible sets of personal weblogs. Its 143 entry and 137 comment sections have been allocated to two separate subcorpora: subcorpus A for entry sections and subcorpus B for comment sections. On the whole, the corpus data yields a total word count of 123,242 tokens and comprises a sufficient collection of weblog data for a quantitative and qualitative analysis of weblog cohesion (cf. table 9). Following the functional and topical selection of appropriate candidates for the AWC, it was decided that the corpus should be divided into two main sub-corpora. One of these sub-corpora contained the entries of all weblogs and the other one included their comments. The two sub-corpora were named A for the entries and B for the comments. Both corpora A and B consisted of further subsets of data of which there were four types:<sup>63</sup>

1. raw data retrieved from the Internet (including words, pictures and hyperlinks)
2. pre-arranged text data in form of word tables, classified according to principal criteria, e.g. word count, number, position and size of pictures, number and type of hyperlink, etc.

---

<sup>62</sup> I shall henceforth refer to this corpus using the acronym AWC (Augsburg Weblog Corpus).

<sup>63</sup> All of the following data sets can be reviewed in more detail in the appendix of this study.

3. P(arts)O(f)S(peech)-tagged verbal text data stored in individual word files<sup>64</sup>
4. Result sheets for verbal cohesion.

### Augsburg Weblog Corpus (AWC)

Sub-Corpus A: Entries		Sub-Corpus B: Comments	
Weblog Title	Length	Weblog Title	Length
All Adither	4270	All Adither	7631
Back Home Again	5254	Back Home Again	4356
Beetle's Memories and Ravings	4429	Beetle's Memories and Ravings	8281
Blog of a Good Time	9289	Blog of a Good Time	3403
Five Blondes	4994	Five Blondes	5486
Mushy's Moochings	3446	Mushy's Moochings	3428
Oh my Words!	8694	Oh my Words!	6646
Shellie's Sentiments	6553	Shellie's Sentiments	4988
Whatever I think	6991	Whatever I think	17206
NonSoccerMom	10469	NonSoccerMom	3728
$\Sigma$	58089	$\Sigma$	65.153

$\Sigma$  TOTAL = 123,242 words

Figure 9: The Augsburg Weblog Corpus (AWC)

While methods were taken to ensure the functional and topical consistency of the weblog data, it proved near to impossible to guarantee the stability of other (socio-linguistic) variables. An attempt was made to elicit crucial socio-linguistic information about the bloggers via the digital questionnaire. Most bloggers were willing to reveal basic personal sociolinguistic clues about themselves, including their age, sex, nationality, profession or place of residence. If blogger's personal information could not be retrieved "directly" through the questionnaire, the "about-me" sections of weblogs were consulted to fill in the information as best as possible. The sociolinguistic information either derived through the questionnaire or through the "about-me" section of a weblog might be false or misleading. We simply have no way

<sup>64</sup> POS tagged data sheets represent verbal data which have been parsed and tagged with the help of the parsing tool *tree tagger*, which is available for free on the Internet: <http://Internet.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/> (11/27/09)

of knowing to what extent Internet identities actually reflect the nature of those who composed them. The information gained in this process remains a reasonable stepping stone for the context-sensitive interpretation of this study's empirical results. The following represents a comprehensive list of the bloggers and their social backgrounds:

<i>Weblog</i>	<i>Age</i>	<i>Sex</i>	<i>Nationality</i>	<i>Residence</i>	<i>Profession</i>
Five Blondes (Collaborative Blog)	20 – 30	Female	Canadian	London (UK) Windsor (UK) Ontario (Canada) Calgary (Canada) Ottawa (Canada)	Student, Students, Social Worker
Back Home Again	20 – 30	Male	US- American	Indiana (USA)	Computer Scientist
All Adither	30 – 40	Female	US- American	Seattle (USA)	Freelance Writer
Beetle's Memories and Ravings	?	Female	British	Yorkshire (USA)	Graphic Designer
Mushy's Moochings	60 – 70	Male	US- American	East Tennessee (USA)	Retired (Telecommunication)
NonSoccerMom	30 – 40	Female	US- American	Texas (USA)	University (Physics department)
Oh my Word	40 – 50	Female	US- American	Paris (France)	Film Actor, Writer
Shelli's Sentiments	40 – 50	Female	US- American	Minneapolis (USA)	Nurse
Whatever I think	50 – 60	Female	US- American	Martha's Vineyard (USA)	Retired (Health Food Store Owner)
Blog of a Good Time	20 – 30	Female	US- American	Pittsburgh (USA)	Graduate Student

Most bloggers are born and live in different parts of the United States of America. One weblog is written collaboratively by five sisters. They reside in different parts of Canada and Great Britain. All other weblogs are only written by one single blogger. It



will thus be interesting to compare the cohesive results of the single-author weblogs to the one which is co-written by multiple bloggers.

The professional background of the bloggers is mixed. It is interesting to note that at least half of them have had some sort of higher education. Bloggers' professions vary considerably, ranging from freelance writers and computer scientists to retired store owners and film actors. In addition, most bloggers are women and only two weblogs are written by men. This does not suggest, however, that personal weblogs are typically written by women (cf. chapter two). The selection of female bloggers was totally random and the selection of other personal weblogs might quite reasonably have resulted in a different picture. With respect to age, the majority of bloggers are between thirty or forty years old, only few are considerably younger or older. No weblogs written by teenagers were included in the AWC.

Following this first socio-linguistic profile of the bloggers whose material has been used for the corpus, I shall now proceed to introduce some further methodological steps which I took to make the corpus even more accessible for the automatic and manual analysis of cohesion.

## **5.2 The Segmentation of the Data**

The raw data for the AWC was excerpted from the entry and comment sections of the ten weblogs in a fixed time period between 1<sup>st</sup> July and 31<sup>st</sup> July 2007. It was first copied and pasted into basic data tables, which were to act as a springboard for all of the ensuing steps of the empirical study. Each entry and comment was then labelled individually with the help of coded identification tags; one tag including three very basic but recognizable pieces of information:

- the weblog to which the entry/comment belonged (acronyms)
- the date of publication
- the title of the entry (acronyms).

The id-code *AA/020708/ens* would, for instance, refer to the title of the weblog (here: AA: All Adither), followed by the date of publication of the respective entry / comment (020708: July 2<sup>nd</sup> 2008) and finally revealing the title of the entry / comment (ENS: Earth's new soundtrack). The purpose of this measure was to facilitate a later identification and easy localisation of individual entry or comment units both within the corpus as well as in this text.

As a next step, each entry and comment section was prefaced and indexed by a outlook page starting each individual set of data. This page served navigation and orientation goals. It instantly presented each entry or comment which was included in the set and revealed their relative position within the file. The outlook pages were followed by data tables. Each table comprised eight different columns: two of which provided most basic information about each entry or comment, i.e. its identification code and verbal content.<sup>65</sup> The remaining columns on the right-hand side added more enhanced information on the units:

- the number of words (word count)
- the most frequent words (theme tags)
- the type and kind of pictures used (icons)
- size and position of pictures (size/pos)
- number and verbal anchor of hyperlinks (hyperlinks (no) and hyperlinks (co)).

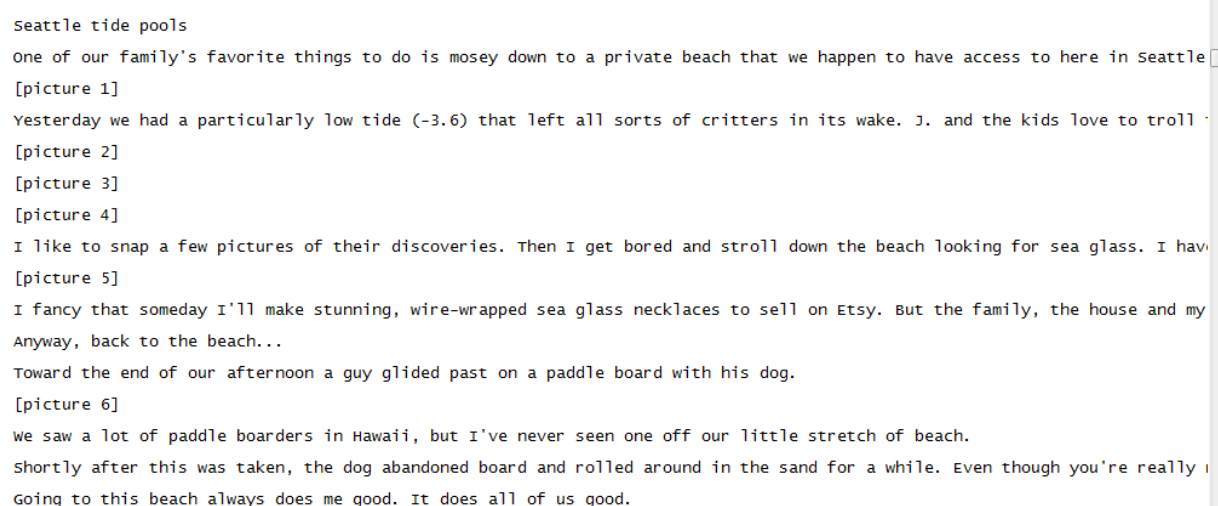
Entry 8	Main Content	Theme Tags	Word Count	Icons	Size/Pos	Hyperlinks (no)	Hyperlinks (co)
AA/160708/wonshpjlotc	Why OK! never should have put Jamie Lynn on their cover	- Jamie Lynn, mom (5) - baby, teen (4) - cover, motherhood, time (3)	510	1. picture (cover of OK magazine)	1. - Size: 200px × 265px= medium - Position: initial, adjacent (left)	1	1. <a href="#">They're out there</a>

Entry 9	Main Content	Theme Tags	Word Count	Icons	Size/Pos	Hyperlinks (no)	Hyperlinks (co)
AA/170708/hteahwyb - tic2.0	How to earn a living with your blog – Tongue in Cheek 2.0	- post (17) - blog (6) - blogger, pictures (4) - site (3)	462	1. picture (photo + text)	1. - Size: 300px × 110px= medium - Position: intermediate	8	1. <a href="#">MommyPie</a> 2. <a href="#">blogger's conference</a> 3. <a href="#">gibberish vernacular</a> 4. <a href="#">snarky commentary</a> 5. <a href="#">children</a> 6. <a href="#">livestock</a> 7. <a href="#">butter and red meat</a> 8. <a href="#">cool links</a>

Figure 24: Data Tables of Weblog Entries with Enhanced Information Columns

<sup>65</sup> Conveniently, the “main content” column of the table was designed to be extendable and retractable with a click of the mouse. This feature proved particularly valuable in the comparison of background information of different entries or comments units.

Evidently, the use of the computer-based tools made it necessary to transfer the raw data into specific data files first (e.g. txt.files and cfg.files). This required certain modifications to the raw data excerpted from the Internet. For instance, the text files had to feature verbal content only to be processed by the automatic tools. Consequently, all hyperlinks and pictures present in the raw data had to be deleted manually and replaced by specific verbal substitutes in the files, i.e. *[hyperlink]* and *[picture]*. Figure 24 shows one such converted text file prepared for automatic text analysis. The next methodological stage comprised the use of two main computer-based tools: the automatic concordancer (*antconc*®) and the POS-parser (*tree tagger*®). Both applications are freely available on the Internet. Based on the search of cohesive trigger words, the program *antconc* was applied to generate concordance lines of potential cohesive relations. The search for these words as well as the subsequent classification of concordances greatly improved on the manual detection and quantification of grammatical cohesive relations. The tree tagger software, on the other hand, was used to facilitate the detection of lexical cohesion. It automatically identified and classified the parts of speech in all weblog entries and comments. Once the POS tags were encoded, nouns, verbs and adjective could be indexed with the help of different colour codes. Both the POS tagging and the visualization technique contributed considerably to the later manual identification and classification of lexical cohesion.



seattle tide pools

one of our family's favorite things to do is mosey down to a private beach that we happen to have access to here in Seattle

[picture 1]

Yesterday we had a particularly low tide (-3.6) that left all sorts of critters in its wake. J. and the kids love to troll

[picture 2]

[picture 3]

[picture 4]

I like to snap a few pictures of their discoveries. Then I get bored and stroll down the beach looking for sea glass. I have

[picture 5]

I fancy that someday I'll make stunning, wire-wrapped sea glass necklaces to sell on Etsy. But the family, the house and my

Anyway, back to the beach...

Toward the end of our afternoon a guy glided past on a paddle board with his dog.

[picture 6]

We saw a lot of paddle boarders in Hawaii, but I've never seen one off our little stretch of beach.

Shortly after this was taken, the dog abandoned board and rolled around in the sand for a while. Even though you're really

going to this beach always does me good. It does all of us good.

Figure 24: A Text File Prepared for Automatic Text Analysis

While the search for grammatical cohesion could be facilitated by the search for specific cohesive forms with the concordancer, this technique could not be pursued

with lexical cohesion. The latter simply did not comprise clear-cut formal indicators which could be searched with concordance tools.

While grammatical cohesion primarily circled around a closed set of searchable cohesive forms, lexical cohesion did not comprise such a list of searchable items. In fact, on a formal plane, lexical cohesion can relate to all kinds of open-class items (e.g. noun phrases, verb phrases and adjective phrases, etc.). Much more than grammatical cohesion, its relations are based on semantic grounds rather than formal cues.

I had to take a different route toward analyzing lexical cohesion which exceeded pure manual labour. For this purpose, the raw text data again had to be saved as computable cfg.files which could then be processed by the tagger program. The linguistic data was automatically parsed and tagged by the program. In fact, the software regrouped all tokens in vertical order, attaching coded word class labels on the horizontal plane of the page.

In	IN	in	
between	IN	between	
laughing	VVG	laugh	
at	IN	at	
her	PP\$	her	
lameness	NN	lameness	
I	PP	I	
almost	RB	almost	
fell	VVD	fall	
asleep	RB	asleep	
because	IN	because	
she	PP	she	
is	VBZ	be	
so	RB	so	
boring	JJ	boring	
.	SENT	.	
She	PP	she	
has	VHZ	have	
like	IN	like	
one	CD	one	
line	NN	line	
in	IN	in	
that	PR	that	

Figure 25: A Parsed and Tagged Text File with Different Colours for Quick Identification Routines

In figure 25, we can see that each lexical item is listed along a vertical trajectory. In addition, it is attached to a word class tag adjacent to it, on the horizontal plane. Following on the far right is the base which the program has used to identify the lexical item and recognize its corresponding word class.

The visualization of the colour codes attached to nouns, verbs and adjectives was conducted manually with the help of the “search and tag” application in Microsoft

word files. The colours used in this process were to represent the following main word classes:

1. yellow tags: proper nouns (NP) or common noun (NN); both singular and plural
2. green tags: adjective (JJ)
3. blue tags: full verbs (VV), forms of the verb “to be” (VB), forms of the verb “to have” (VH).

The colour codes allowed me to recognise noun phrases, verb phrases and adjective phrases across clauses more efficiently. It equally accelerated the identification of possible lexical cohesion in the manual part of the analysis.

The colour code should not suggest, however, that only noun, verb and adjective phrases were examined for potential cohesive relations. In principle, all lexical items were scrutinized. Besides phrases, pronouns or conjunctions, for instance, could recur between clauses and thus install further cohesive relations. Different realizations of pronouns or determiners (in subjective or objective case), such as *me*, *my*, *mine* or *he*, *his*, *him* were also evaluated and formed a crucial part of the network of lexical cohesion in weblogs.

Toward the identification of grammatical and lexical cohesive relations, I needed demarcate their scope in more detail. In other words, it became necessary to distinguish between intraclausal and interclausal cohesive relations and select whether both types of relations were considered part of this study. On a methodological plane, clauses were simply distanced by added spacing in the text files. The additional visual separation of clauses thus made possible a more convenient and co(n)text-sensitive distinction between intra- and interclausal cohesive relations.

### **5.3 The Manual Analysis and Final Evaluation of the Data**

I shall now return to the manual examination and analysis of the data sets following their automatic annotation. First, we will turn to the analysis of grammatical cohesion on the basis of automatic text concordances. The manual analysis of the concordance lines comprised four crucial steps:

1. Each concordance hit had to be assessed with respect to the cohesive or non-cohesive usage of the respective search term. Non-cohesive hits were deleted

from the files and labelled “not cohesive”, usually accompanied by the particular non-cohesive function of the item in context. Note again, at this stage, that intersentential and interclausal relations were considered cohesive, while intraclausal relations were excluded from the study.

2. If concordance hits included a search term which referred to (textual, situational or cognitive) domains external to the text units, these relations were termed exophoric and marked accordingly in the files.
3. Endophoric cohesive bonds were classified as either anaphoric or cataphoric relations. The respective referent of cohesive cues was identified in the context of the utterance and added to the concordance line indicated by an arrow.
4. In contrast to entry sections, comment sections additionally featured cohesive relations between the contributions of different users. There were relations
  - within the comment
  - between the comment and its related entry
  - between different comments.

It was concluded that these differences should equally be indicated in the text files. Therefore, cohesive relations within comments were tagged as either *anaphoric* or *cataphoric*, depending on their respective direction of reference. Additionally, relations between comments and entries were identified as *entry* and relations between comments as *cross-comment*.

The POS-tagged text files equally needed additional scrutiny with respect to potential misclassifications resulting from the automatic tagging procedure. There are three main classificatory problems:

- some lexical items were neglected in the tagging process
- some lexical items were assigned false word classes
- some lexical items were misplaced in the tagging process.

These three problems could only be resolved by a secondary manual evaluation and correction of the data sets. Finally, the interclausal cohesive relations were analysed

in detail. The results of the analysis then needed to be transferred to a novel set of result sheets which specified both the lexical ties involved in lexical cohesion as well as the underlying cohesive category (cf. figure 26). A further complexity surfaced in the analysis of lexical cohesion in and between weblog comments. In order to elicit the polyphonic structure of interaction in comment sections, I had to distinguish between cohesive relations within and between comments. As we can see in figure 26, the last column of the result sheet specifies the cohesive categories of the relations found within the text. The relation is tagged “IC” for “inside comment” and “CC” for “cross-comment” revealing the scope of reference exclusively for the interactive discourse setting of weblog comments. Note that the third column (“comment (co)”) contains several different boxes. Each one includes the verbal text of a single comment which aligns with an underlying entry, that can be identified by the first column (“entry”).

Entry 1	Comments (no)	Comments (co)	Noun Phrase	Lexical Relation
BHA/010708/tb2008	7			
		Sounds like you had a blast!		
		I'm taller than the Kerrville hills. It was great seeing you again and great recap!	you – you	Total Recurrence (CC)
		Terrific recap. You captured it perfectly.	Recap – recap Recap – captured You – you	Total Recurrence (CC) Partial Recurrence (IC) Total Recurrence (CC)
		thanks, <u>darlin</u> . <u>you</u> know you are welcome anytime. <u>plus</u> , i like watching you cut fruit. <u>and</u> cut it. <u>and</u> cut it. Chop <u>chop</u> , my monkey man!  for real, thanks for coming. <u>and</u> special thanks for getting tlg and <u>erica</u> there. <u>can't</u> wait to see ya'll again.  <u>be</u> good, <u>i</u> mean it.	You – you You – you You – you Cut – cut Cut – cut Cut – chop Cut – chop Thanks – thanks Welcome- coming You – you	Total Recurrence (CC) Total Recurrence (CC)  Total Recurrence (IC) Total Recurrence (IC) Total Recurrence (IC) Co-Hyponymy (IC) Co-Hyponymy (IC) Total Recurrence (IC) Partial Recurrence (IC) Total Recurrence (IC)

Figure 26: Tables for the Manual Analysis of Lexical Cohesion (here: Comments)

The results of the search for grammatical cohesion in the corpus as well as the results of the analysis of lexical cohesion were then counted and classified according to the cohesive categories to which they belonged. Still, these results could not be compared to each other because the text length of the individual files varied considerably. Therefore, the results needed to be normalized to a frequency of 1000

words. I choose to draw on a regular standardization procedure introduced by Biber (1988:14). I divided the total number of cohesive relations for each category by the total length of the respective file and multiplied the resulting number by 1000. The procedure produced the average number of cohesive pairs per 1000 words for the particular cohesive category in the specific text file used. The resulting numbers now figured as *normalised frequencies* so that their quantities could be compared to each other as well as to other quantitative results of similar studies of cohesion. The corpus was now ready to be evaluated and interpreted, and this is what I shall do presently in the two following chapters.

#### **5.4 Preliminary Methodological Reflections**

The following chapter will discuss the results of the empirical analysis of grammatical cohesion in the Augsburg Weblog Corpus (AWC). The main part presents and interprets the quantitative distribution of *personal*, *demonstrative* and *comparative reference* in the weblog entries and comments. I will close with a description of cohesive variation of *substitution*, *ellipsis* and *conjunction* in the AWC. More specifically, this comparison will focus on the frequency, the distribution and the types of reference, following three main objectives:

- (1) to compare and discuss the quantitative distribution of grammatical cohesion in the weblog entries and comments
- (2) to compare the use of grammatical cohesion in the weblogs with its use in prototypical spoken and written text genres
- (3) to use the findings obtained in (1) and (2) to form a preliminary hypothesis on the monologic or dialogic character of personal weblogs.

The empirical results will be interpreted by recourse to the sociolinguistic information (e.g. status, profession, age, gender, etc.) and media-related exigencies of weblog communication (e.g. character limitations, use of semiotic modes, time and space restrictions, etc.). All data frequencies will henceforth be represented in normalized frequencies unless indicated otherwise. This is to assure the general comparability of language data within weblogs as well as between weblog entries and comments. While the analysis and interpretation will be primarily based on quantitative results, often, I will interpret the latter by recourse to qualitative evaluations of individual examples, which are drawn from the weblog corpus itself. As we shall see, it is



necessary to put the quantitative facts on a reliable qualitative footing to account for its ensuing interpretation.

Finally, a preliminary note of caution seems to be in order. The following two chapters aim to pinpoint the personal weblog between two communicative extremes, i.e. the spoken, spontaneous face-to-face conversation and the written, formal academic articles; as it is these two extremes which linguists usually refer to when they wish to talk of dialogue or monologue. I am fully aware of the fact that the investigation of spoken and written language cannot only be understood by contrasting these two extreme forms of interaction; yet they conveniently serve as a blueprint for the interactional localisation of the weblog. The aim of this study cannot consist of a clear-cut identification of the personal weblog with either ends of the interactional divide. It will be much more interesting to look at the individual features which make the personal weblog lean towards either side of the spectrum. The monologic, dialogic (or indeed hybrid) character of the personal weblog will thus be based on the evaluation of its monologic or dialogic cohesive texture in and across the AWC weblogs.

It is the cohesive texture of personal weblogs and its relation to the interactive dimension of weblog communication, which has so far been unexplored in linguistic research. Previous studies on spoken and written language have generally based their interpretations on word frequency counts of linguistic forms or constructions in various spoken or written text corpora (cf. Drieman 1962, Chafe 1982, Altenberg 1986). At first, studies based their understanding of spoken and written language on a dichotomous distinction of two prototypes only, i.e. two-party conversation for spoken language and academic articles for written language. As a result, some recurring formal differences between these two communicational extremes emerged (e.g. clause complexity, conjunctions, attributive adjectives, interpersonal pronouns, etc.). However, more recent studies have proven that such crude separations of spoken and written texts based on bundles of positivistic data are inconclusive for other spoken and written text genres.<sup>66</sup>

Biber (1988, 1995) therefore proposes to forsake the original binary view of spoken and written language and replace it with a dimensional perspective on

---

<sup>66</sup> For instance, Beaman's work on clause coordination and subordination (1984) suggests that high frequencies for complex clauses in academic prose may not hold for other text genres like recipes or print advertisements.

typological classification. He draws up five different dimensions. These answer to different functional orientations of the text:

1. *involved vs. informational orientation*
2. *narrative vs. nonnarrative concerns*
3. *explicit vs. situation dependent reference*
4. *overt expression of persuasion*
5. *abstract vs. nonabstract style*

(Biber 1988:8)

As we can see, Biber's textual dimensions are principally governed by functional (purpose-driven) criteria. Biber now assumes that each dimension is defined by prototypical linguistic forms. Individual texts can thus be positioned on a scale of generic affirmation of these dimensions identified through the quantitative recognition of formal means. This view is indeed superior to a binary categorisation of spoken and written language. In this study, however, I am not interested in the classification of personal weblogs as a weblog genre. Rather, the analysis of cohesive distributions in and across weblogs should yield findings which pinpoint the communicative status of weblogs between spoken dialogues and written monologues. To this end, the interpretation will be based on frequency counts of discourse relations rather than the elicitation of discourse forms only. This additional methodological step enables us to elicit the actual interactive dimension of weblogs as mirrored by the texture they assume. Following recent work by Tanskanen (2006), this study does not simply tally word frequencies for lexical constructions but, in addition, assesses the role of these forms as they enter into interclausal cohesive relationships. As a result, the underlying cohesive texture of personal weblogs can be compared. The analysis promises to shed new light on the construction and interactive genesis of cohesion in personal weblogs. In the end, the quantitative examination of cohesion in personal weblogs produces a more profound understanding of the monologic or dialogic nature of this pervasive weblog genre.

## The Analysis

### Part I: Grammatical Cohesion

---

#### 6.1 Reference in Weblog Entries

I start the investigation of weblogs' cohesive texture with a discussion of personal and demonstrative reference in weblog entries. The following table presents the distribution of endophoric and exophoric discourse relations for personal and demonstrative reference in and across the AWC weblog entries.

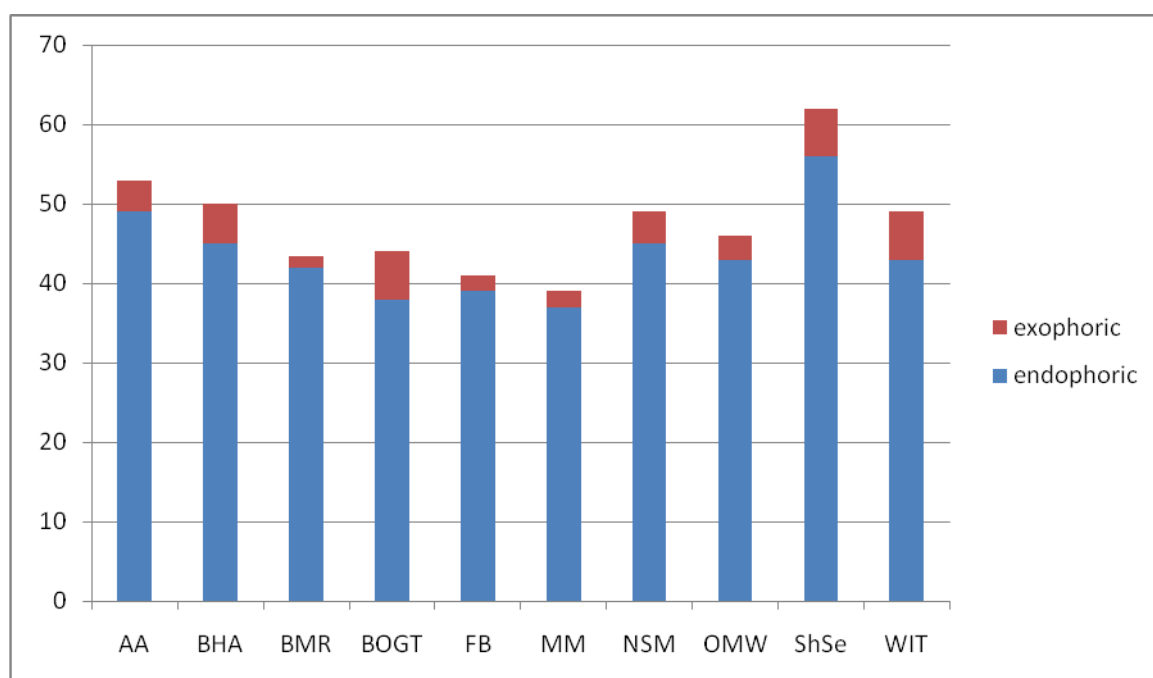


Table 10: Endophoric vs. Exophoric Relations in Weblog Entries

The majority of reference relations in the AWC entry corpus is endophoric. Exophoric relations are very infrequent in the weblog entries. In table 11, endophoric relations are further divided into personal and demonstrative reference. The columns indicate that personal reference clearly dominates over demonstrative reference. Three-fourths of all endophoric relations are personal reference relations; while demonstrative reference is much less represented (cf. table 10).

The high percentage of endophoric reference and the low percentage of exophoric reference in the weblog entries align with the discursive profile of the classical written monologue. As discussed in Biber (1992:234), text genres differ in

the extent that they make use of endophoric and exophoric references. The following endophoric and exophoric dimensions can be aligned with nine different spoken and written text genres:

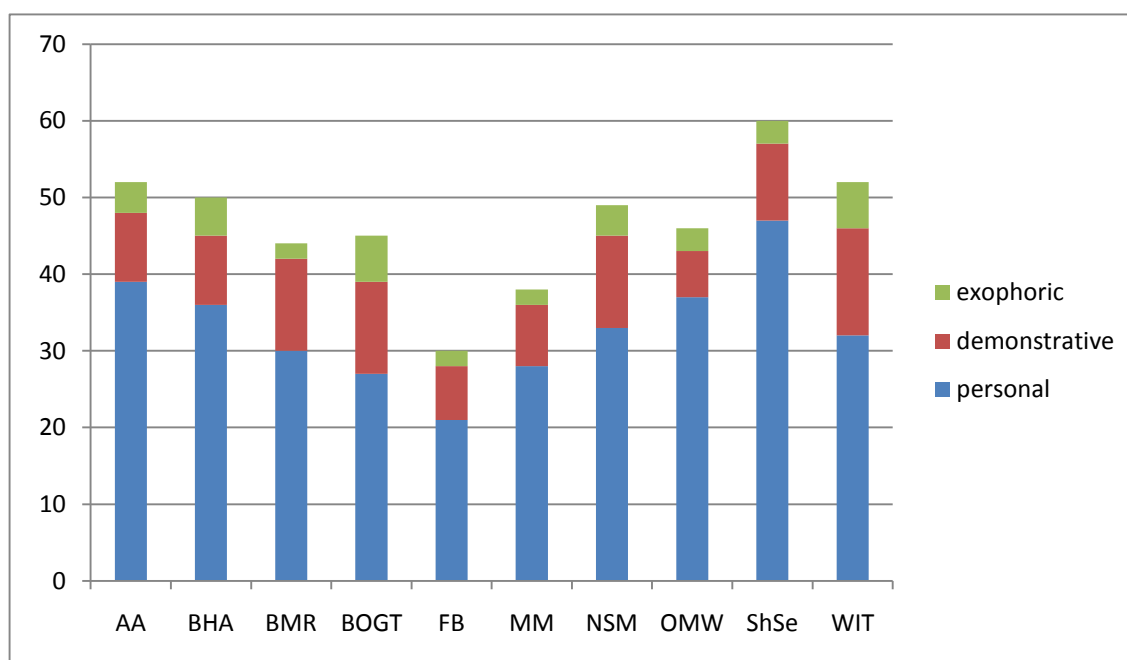


Table 11: Personal and Demonstrative Reference in Weblog Entries

In table 11, Biber (1992:235) presents only normalized frequencies for pronouns. His chart includes only *anaphoric*, *exophoric* and *vague pronouns*.<sup>67</sup> He notably excludes cataphoric pronouns from his investigation. As cataphoric pronouns only occur in very limited both in Biber's corpus and the AWC, neglecting this sub-category has no actual effect on the quantitative results of both analyses. However, Biber includes interpersonal pronouns, i.e. *I*, *you* and *we* in his category of exophoric pronouns. In this study, I have chosen to classify these as instances of *lexical cohesion*. This methodological difference is responsible for some quantitative differences between both studies. I will now consider these differences more carefully. In table 12, we can see that the frequencies for exophoric pronouns in Biber's study are generally high for spoken genres and low for written genres. This finding can be explained by the interpersonal pronouns, which are more prominently used in spoken than in written discourse.

<sup>67</sup> Biber (1992:220) defines *vague pronouns* as pronouns which relate to an "action or stretch of discourse" rather than to a "previous referring expression" (Biber 1992:220).

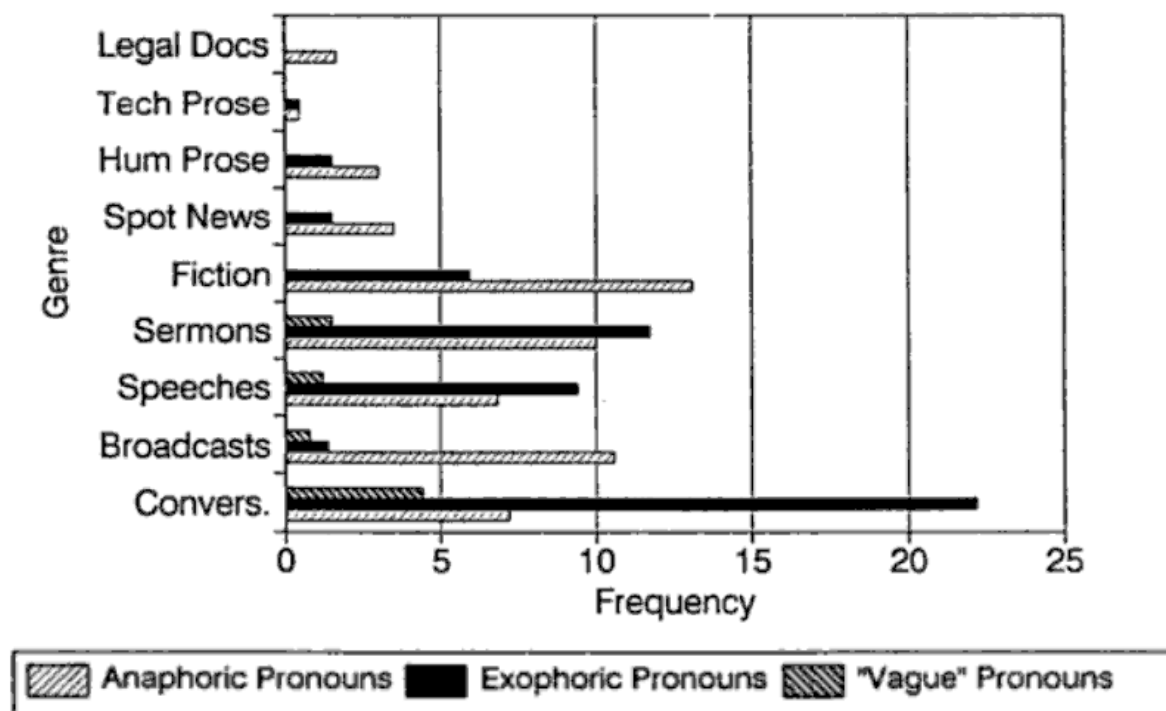


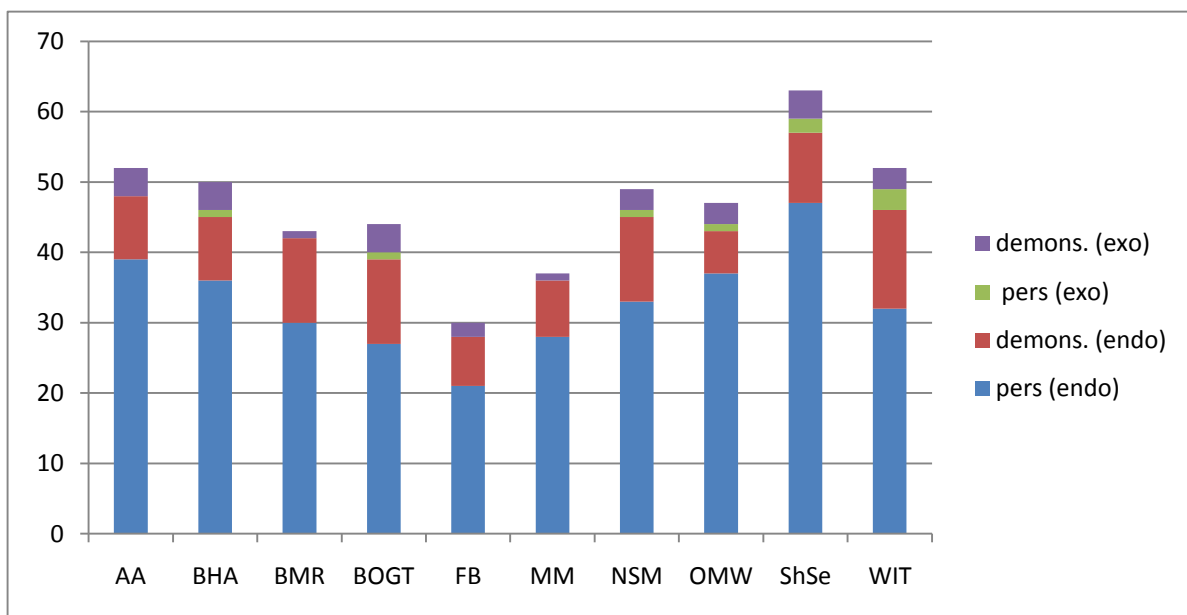
Table 12: Cross-Generic Variation of Anaphoric and Exophoric Pronouns (Biber 1992:235)

Conversations usually comprise constant changes of speaker and hearer roles. These roles are being re-negotiated and re-adjusted to suit the interactive demands of the underlying discourse. This is reflected by the frequent use of interpersonal pronouns in conversational text genres. In written monologic discourse, the interactional setting is much less demanding. Here, speaker change is not an issue at all, and interpersonal pronouns are consequently applied less frequently. Biber reflects on this difference between spoken and written discourse in his findings on exophoric pronouns: most spoken text genres, except for broadcasts, exhibit high frequencies for exophoric reference while written genres show comparatively lower ones. If we compare the two prototypes of monologic and dialogic interaction, i.e. *humanities academic prose* and *conversation*, this difference becomes most evident.

Furthermore, we can contrast the findings for the use of anaphoric and exophoric reference in weblog entries with Biber's results. The following picture arises: the profile of exophoric reference in the AWC weblog entries is more readily aligned with the genre *humanities academic prose (HAP)* than with the genre of

*conversation (CON)*.<sup>68</sup> I suggest that this is a clear indication of weblog entries being more similar to the referential profile of written monologues than the one revealed for spoken dialogues. Note that the percentage of *HAP* shows a higher ratio for exophoric reference than the one for weblog entries. Nonetheless, this difference is quantitatively low. If all interpersonal pronouns were deleted from the list of exophoric pronouns in the *HAP*, the cohesive profiles of both *HAP* and weblog entries can be expected to converge considerably.

Next, I shall investigate the quantitative distribution of different types of reference within and across the different AWC weblog entries. To this end, I first examine how many of the endophoric relations in weblog entries are classified instances of personal or demonstrative reference. The following two tables provide a first general overview on personal and demonstrative cohesion in the entry sections of all AWC weblogs:



	pers (endo)	demons. (endo)	pers (exo)	demons. (exo)
AA	39	9	0	4
BHA	36	9	1	4
BMR	30	12	0	1
BOGT	27	12	1	4
FB	21	7	0	2
MM	28	8	0	1
NSM	33	12	1	3
OMW	37	6	1	3

<sup>68</sup> The case of interpersonal pronouns as a subtype of reiteration in lexical cohesion will be discussed in the next chapter which will elaborate on the individual role of interpersonal pronouns in personal weblogs.

ShSe	47	10	2	4
WIT	32	14	3	3

Tables 13 and 14: Personal and Demonstrative Reference in Weblog Entries

Note that most exophoric relations are demonstrative and only few are personal. The dominance of demonstrative exophoric references reflects blogger's frequent use of *time deixis*. In the weblogs, time deixis mostly occurs in the form of demonstrative determiners modifying one time-related noun phrase, for instance *this summer*, *this week*, *that day*, etc.. The expressions refer to specific times or dates which cannot be resolved by endophoric reference. In conversational settings, time deixis is commonly used to refer to the actual time in which the ongoing discourse takes shape. Interlocutors thus need to be aware of the time and place of speaking to make sense of deictic discourse elements. Weblog authors and users do not share a common discourse situation. Each interlocutor literally sits in front of a different computer at different places and times. So, the common discourse situation has to be virtualised. It has to be compensated by the virtual setting of the computer screen. Hence, users commonly use time deixis if they are sure that others can easily resolve the reference by recourse to weblog elements surrounding the actual text units. One of these elements is the so-called *time stamp*.<sup>69</sup>



Figure 27: Time Stamps and Date Indicators in Weblog Entries ([www.alladither.com](http://www.alladither.com))

<sup>69</sup> A time stamp is an automatic indicator of the time at which a given weblog entry was published on the Internet.

Time stamps are automatically attached to each entry. They allow users to identify the exact time of day at which the weblog entry was published. The latter can use the temporal information made available through time stamps to find out whether a weblog entry was published in the morning, afternoon or at night. This may help users to identify some types of exophoric references. More commonly, time deictic references in weblog entries point to specific week days or months. These temporal pieces of information are not provided by the time stamps. Rather they have to be indicated by some other exophoric element of the weblog, called *date indicators*. It appears that most bloggers start their entries with an explicit mention of the date at which the entry was written. It is then usually succeeded by the entry title. *Date indicators* are the most important exophoric resource for the identification of *time deixis* in weblog entries and comments. While time stamp inform users about the exact hour at which a contribution was published, date indicators locate the exact day, month and year of its creation.

Both time stamps and date indications are called *exophoric elements* of the weblog because users find them placed externally of the entry or comment text units. The relation between demonstrative exophoric reference and time stamp or date indicator is classified as an exophoric rather than an endophoric relation. See below for a list of demonstrative exophoric references in the weblog *Five Blondes*:

- (42) and I spent **this past week** at my best friend Michelle's cottage in Muskoka.  
(FB/110708/rmaamlvfb)
  
- (43) he will be growing his beard back starting **this weekend!** I love the facial hair.  
(FB/090708/dtpaoswd)
  
- (44) **This week** I decided it was probably because, while I swam in lakes and oceans tons as a child, I didn't spend any time on boats (besides canoes) [...]  
(FB/110708/rmaamlvfb)
  
- (45) Kent is naturally good at every sport he tries, he even tried waterskiing with one ski **this week**, [...]  
(FB/110708/rmaamlvfb)



Most exophoric references in the weblog entries were found to be instances of demonstrative rather than personal reference. We can expect this finding to reflect the frequent use of time deixis in weblog entries. In fact, the majority of demonstrative exophoric references were triggered by the demonstrative determiners (*this* and *that*) and not by demonstrative pronouns; a finding which adds up to this line of interpretation.

Let us now consider the personal reference relations in weblog entries. The high ratio of personal reference in table 13 can be explained with respect to the mode of production which governs the composition of weblog entries. The latter are usually written by a single person who has ample time to plan, compose and structure his contribution in advance, before he publishes it online. Entry writing is thus characterized by a high degree of compositional control on the part of the blogger. There is only a one-sided interaction in weblog entries; bloggers themselves can write entries but users can only engage in comment sections. They cannot publish entries unless they have acquired additional publication rights from the author of the weblog. Entries are thus characterized by one-sided interaction only. In fact, even though users may write comments to a given entry, the ensuing interaction takes place in an asynchronous manner. Entries also regularly include quite elaborate narrative descriptions of events or complex verbal renditions of feelings, reflections or experiences. As a result, entry word length as well as sentence complexity is considerable. In other words, short entries have been found to be the exception rather than the rule. This fact is one of the more elementary differences between weblog entries and comments; the latter being considerably shorter than their respective entries. However, the salience of personal reference equally suggests something else: cohesive ties can be expected to hold across multiple sentences; even across the entire entry, and thus form veritable chains of identity forged by the recurring use of personal pronouns or determiners which progress discourse agents throughout the continuing discourse. The concordance, which is illustrated in table 28, is drawn from the weblog *All Adither* and serves as a case in point. We can see that there are two identity chains in the concordance excerpt: *she* → *white stay at home mom* and *she* → *Jamie Lynn*. Both of them exceed the relation of a cohesive tie which only comprises two contiguous interclausal elements. Furthermore, the chains comprise not two but eight segments which progress topical agents through the ongoing discourse. Chaining tendencies are especially strong in weblog entries.

Every entry in the AWC entry corpus featured at least one identity chain with an average length of four to ten interclausal elements. Most exhibited multiple chains at once. Previous studies have shown that the elicitation of identity chains (based on grammatical reference) and similarity chains (based on lexical reference) yield salient discourse agents and topics (cf. Hasan 1984, Hoey 2005). Accordingly, the identification of identity chains in the AWC weblog entries equally produced an overview of their central discourse agents. See below for two additional concordance excerpts from the weblogs *All Adither* and *Five Blondes*:

1	just THIS particular white, stay-at-home mom: • She --> white stay at home mom
2	and asks for help. No drinking, he says. Anything she --> white stay at home mom
3	s. Anything she could offer would be appreciated. She --> white stay at home mom
4	, feels a moment of poignant regret, then decides she --> white stay at home mom
5	regret, then decides she will blog about it once she --> white stay at home mom
6	er internet friends about the poignant regret. • She --> white stay at home mom
7	d Social Engineering for anyone who's interested) She --> white stay at home mom
8	eyes were too desperate. • Later that afternoon, she --> white stay at home mom
9	at, my three-year-old daughter, talks like Linus. She --> Kitty Cat
10	ay of forming her words and I treasure it. I know she --> Kitty Cat
11	. "Being a mom is the best feeling in the world!" she --> Jamie Lynn
12	he tells us. It's all so wonderful and great! And she --> Jamie Lynn
13	y fake-euphoric photos and "interview". Granted, she --> Jamie Lynn
14	aising her baby in one half of a sagging duplex. She --> Jamie Lynn
15	e sister and think, Look! Jamie Lynn is doing it. She --> Jamie Lynn
16	has a great house and her boyfriend proposed and she --> Jamie Lynn
17	ns while her fiance pours her a cup of coffee (is she --> Jamie Lynn
18	even old enough to drink coffee?) and think what she --> Jamie Lynn

Figure 28: Identity Chains in Weblog Entries (Here: All Adither)

1	with talons jumping out a freezer onto my face - they --> talons
2	t, well actually that isn't true, I have been but they --> pictures
3	ichelle's camera. She lost her camera cord and so they --> picture
4	lking to Dad when Dad suggested that we see where they --> dad +?
5	ask anyone who has access to my facebook page and they --> anyone
6	ut if these 50 girls were, in fact, girls - or if they --> girls
7	h donations of money, food, clothing & education. They --> AmmA
[...]	
15	ome days they actually stop working for a bit and they --> Oma & Opa
17	. What do we love about Oma and Opa? We love that they --> Oma & Opa
18	lping us celebrate what's happening in our lives. They --> Oma & Opa
19	our lives. They taught us how to swear in Dutch. They --> Oma & Opa
20	till so active and young that we sometimes forget they --> Oma & Opa
21	s forget they're getting old. We're so proud that they --> Oma & Opa
22	that they have been together for fifty years, and they --> Oma & Opa

Figure 29: Identity Chains in Weblog Entries (Here: Five Blondes)

In the first concordance excerpt, there are three identity chains whose chain length varies considerably (i.e. *she* → *white stay at home mom*; *she* → *Kitty Cat*, *she* → *Jamie Lynn*). The second weblog, *Five Blondes*, is written cooperatively by five different bloggers. It contains three identity chains of considerable length. In the concordance, we can find a long identity chain triggered by the personal pronoun *they*, which co-refers with the preceding noun phrase *Oma and Opa* in the text. These identity chains connect the discursive web of entry sections. However, chains may sometimes skip a number of sentences before the co-reference is resumed by the next personal pronoun or determiner. In most cases, however, identity chains are linked up from one clause to the next, as is shown in the following entry excerpt from the weblog *All Adither*:

When we first had Fruit Bat tested for food allergies, **he** was five months old. **He** had eczema from head to toe and I was desperate to figure out the cause. **He** tested positive for dairy, egg, tree nuts and peanuts. My initial reaction was relief. I could cut those foods from my diet (I was still nursing **him** exclusively)

(AA/280708/tpwigapoofaa)

Hence, identity chains bear witness to the fact that weblog entries develop themes rather extensively in paragraphs. Changes in identity chains usually coincide with changes in the topical structure of the weblog entry. Note that in figure x above the blogger seems to address a number of different issues centred on her new camera (hence the double reference to pictures) and the pictures she plans to upload to her weblog. Then, the blogger starts a new topic related to her grandfather and grandmother (*Opa and Oma*). As a result, a novel strand of identity relations emerges. If we were to go through the individual concordance analyses, we would very likely be able to get some coherent picture of the types of agents used recurrently throughout the entries.

The length of identity chains suggests that topic drifts are uncommon in weblog entries. In fact, rapid topic changes would probably confuse users and are counterproductive to the construal of coherence by users. Bloggers, who usually seek the admiration and affection of a trusted readership, can therefore be expected to avoid topical alternation. Altogether, the high ratio of personal reference manifest in frequent long identity chains is indicative of a monologous discourse setting (cf. Biber 1992). The author needs substantial time and writing space to progress and elaborate his topic conveniently. In contrast, in two-party or three-party dialogues

identity chains do exist, but due to the interactive negotiation of turns, they can be much more constrained in length.<sup>70</sup> In addition, regular face-to-face conversations seem to be less regulated so that topical shifts, frame breaks and referential ambiguities frequently occur. The cohesive profile of personal and demonstrative reference in the AWC weblog entries thus resembles the classic cohesive profile of the monologue.

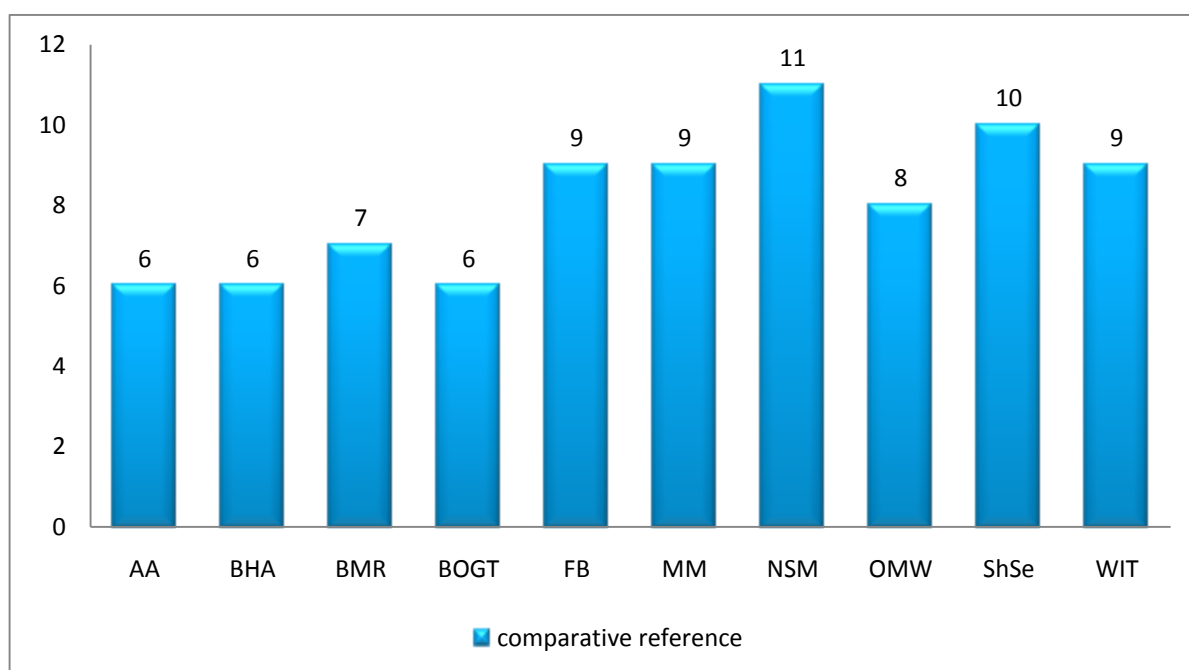


Table 14: Normalized Frequencies for Comparative Reference in Weblog Entries

Table 14 illustrates the normalized frequencies for comparative reference in all AWC weblog entries. The frequencies range from six to eleven with a mean average of eight hits per 1000 words (8 h/w<sup>71</sup>). If we compare comparative reference to the frequencies of endophoric demonstrative reference (10 h/w) and exophoric demonstrative reference (4 h/w), comparative reference claims a middle ground. Personal reference outnumbers comparative reference by four to one while the standardized number of demonstrative reference (i.e. endophoric and exophoric reference) is only marginally higher than comparative reference. Still, table one includes not only interclausal but also a considerable amount of intraclausal relations.

<sup>70</sup> This does, of course, not hold for more specific dialogue settings, such as moderated interviews in which interlocutors are given a lot of time to develop their argument and turn-allocation is regulated by the moderator.

<sup>71</sup> The abbreviation *h/w* will be used henceforth to refer to "hits per thousand words".

If I were to factor out these relations, a slightly different picture would arise with demonstrative reference exceeding the ratio of comparative reference by two to one.

Let us now consider the variation of comparative reference in the different weblogs in more detail now. In table 14, we can basically discern two different categories of weblogs: one has a scope of 6-8 h/w for comparative reference, whereas the other category produces higher ratios of 9-11 h/w. The first class consists of the following weblogs:

- AA: All Adither (6 h/w)
- BHA: Back Home Again (6 h/w)
- BOGT: Blog of a Good Time (6 h/w)
- BMR: Beetle's Memories and Ravings (7 h/w)
- OMW: Oh my word (8 h/w).

The second class assembles the remaining candidates which are:

- MM: Mushy's Moochings (9 h/w)
- FB: Five Blondes (9 h/w)
- WIT: Whatever I think (9 h/w)
- ShSe: Shelli's Sentiments (10 h/w)
- NSM: NonSoccerMom (11 h/w).

The two weblogs which exhibited the highest frequencies for comparative reference are *Shelli's Sentiments* and *NonSoccerMom*. Both exhibit two times the amount of hits per 1000 words compared to the weblogs *All Adither*, *Back Home Again* and *Blog of a Good Time*. Let us compare their individual profiles of comparative reference for each of those blogs more closely now:

	ShSe	NSM	AA	BHA	BOGT
{-er}	2	2	2	1	1
another	1	1	--	1	0
other	2	1	1	1	1
as + (q/a/a)	1	1	2	0	1
more	3	3	1	2	1
less	0	1	--	--	0
equal	0	0	--	--	--
such	--	1	--	1	0
same	1	0	1	0	0
similar	--	--	--	--	0

different	1	0	--	0	0
-----------	---	---	----	---	---

Table 14: Comparative Reference in Five Different Weblog Entries (Normalized Frequencies)

Note that all the weblogs in the table show comparable ratios for the comparative means of:

- the inflectional comparative {-ER} (1-2)
- the lexical item *other* (1-2)
- the lexical item *more* (1-3)
- *as + quantifier / adverb / adjective* (1-3).

The comparative marker *more* is used more frequently in ShSe and NSM, but this difference does not account for the quantitative difference between the normalized frequencies for the first and second class of weblogs previously discussed. It is more likely that the weblogs AA, BHA and BOGT (representing the first class of weblogs), ShSe and NSM (representing the second class) do not differ in terms of frequency but in terms of categorical variation. Note that the ShSe and NSM weblog include a broad range of nine to ten different comparative categories in their entry sections. Interestingly, the frequency of each formal category of comparative reference is not particularly high in these two weblogs. In other words, there is no particular quantitative preference for one or two categories. Rather, the bloggers make consistent use of a broad confection of formal expressions which, in summation, amount to their high ratio of comparative reference.

In the three other weblogs, which represent the second class of weblogs with low general frequencies, the picture is somewhat different. Formal variation in the two weblogs *Back Home Again* (BHA) and *Blog of a Good Time* (BOGT) is low. They only feature five or eight different comparative categories. Although the remaining comparative categories used in these weblogs boost with high ratios, they cannot fully redeem for their lack of formal variety. Consider, for instance, the standardized hits for the form: *as + quantifier / adjective / adverb* in all of the above weblogs. The weblog *All Adither* (AA) may feature more hits than the two weblogs ShSe and NSM but still has a low normalized frequency of comparative reference.

In a similar vein, the weblog *BHA* produced substantial hits in the category *more* but does not show comparable quantities in others. Finally, the weblog BOGT shows sufficient formal variation (it includes ten different formal realizations of the

comparative category). However, it fails to underscore this formal variability with the help of consistent quantities of hits. It therefore falls back in the overall count of hits used for measuring the normalized frequency of each weblog. On the whole, we can follow that the normalized frequencies for comparative reference in weblog entries are not heterogeneous (6-11 h/w). All frequencies are low compared to the frequencies measured for personal and demonstrative reference in the corpus. There are two classes of weblogs. One shows higher frequencies than the other. The difference can be explained by recourse to the formal variation of comparative reference in each weblog. Weblogs which show consistent numbers in different comparative categories have higher ratios than weblogs which exhibit lower formal variation.

Various reasons for these findings can be suggested. The more constrained or more varied use of comparative reference in the different weblogs (class one or two) could be influenced by the topical orientation of the weblog entries. Accordingly, some topics may be more likely to evoke the use of comparative markers than others, e.g. biographical posts, reviews of other people's weblogs or sites, reactions toward bloggers' own previous writings. However, I could not find any conclusive proof for this claim in the two weblogs, which featured high ratios of comparative reference. From a topical point of view, their entries did not deviate significantly from others, and a topical argumentation does not seem to explain the aforementioned findings. Different linguistic proficiencies of the bloggers can equally be ruled out as an explanation for greater formal variability. We may only speculate if bloggers' personal writing experience actually has a profound influence on their linguistic capacity to express themselves more diversely. Unfortunately, to this day, linguistic studies which explore this relationship are missing. In other words, the (socio-)linguistic evolution of language in weblogs is still largely unaccounted for.<sup>72</sup>

To my knowledge, there exists no conclusive quantitative study on the variation of comparative reference across different text genres. We simply have no numerical basis against which we could possibly match the previous findings. However, it is by comparison with the results gained from sub-corpus B of the AWC (comment sections) that we can hope to acquire the necessary frame of reference.

---

<sup>72</sup> One intriguing aspect of diachronic investigations into language change in weblogs is to compare bloggers' individual experience of generic change in the blogosphere with the linguistic change they themselves have undergone over time.

For now, I shall continue describing the frequency variation for the various formal realisations of comparative reference in the weblog entries. In order to gain some perspective, the results of the search for comparative reference in each weblog entry were simply tallied and charted to identify roughly which forms were used most often across all weblog entry sections.

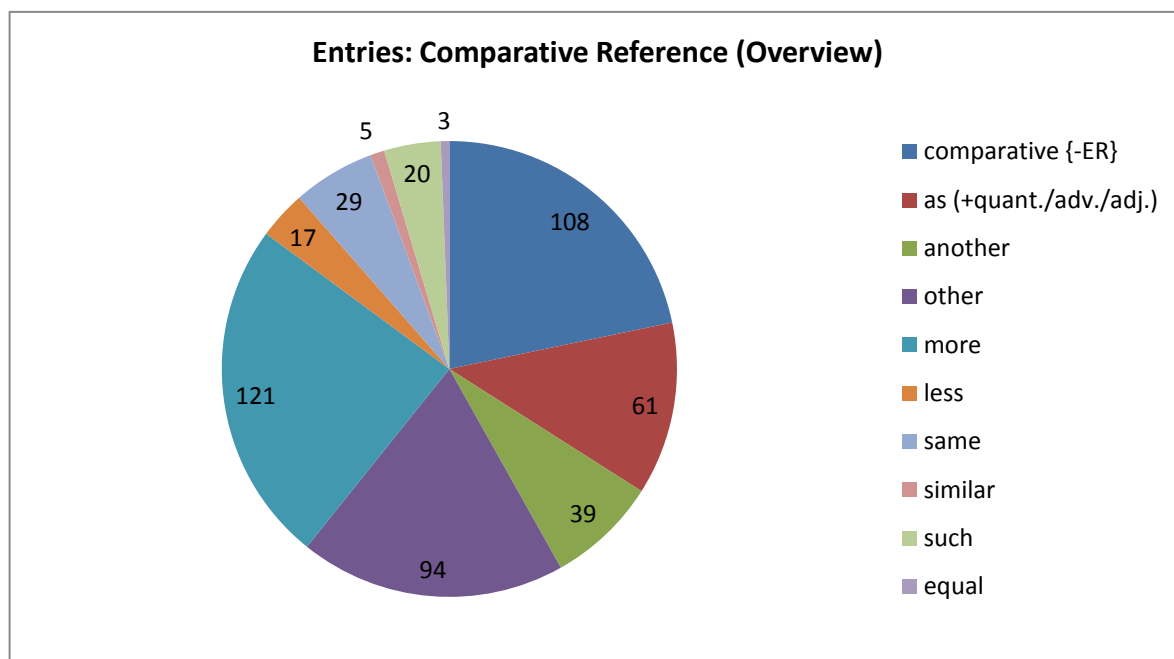


Table 14: Comparative Reference in Weblog Entries (Variation of Cohesive Categories)

What emerged out of this procedure was the following picture (cf. table 15): the largest contributor to comparative reference in weblog entries is the comparative marker *more*. Next is the inflectional suffix {-ER}. It should be noted that the practical identification and/or resolution of cohesive pairs in this category proved to be cumbersome. While few hits could be analysed with little difficulty, other hits proved to be quite challenging. Consider the following example hits in this respect:

(46) And as for N, he seems to be feeling much better, so that's good news.  
(NSM/300708/tpbtybatlos)

(47) [He calls me] Wife. Not The Wife, like his blog terminology. Just like hey, wife, come here for a second. Does this bother me? Not really. It is better than calling me "Woman" I suppose.  
(NSM/280708/syaiihjntt)



In (46), the comparative form *better* connects the preceding implicit assertion of “my husband calls me wife” with the following explicit clause “calling me ‘woman’”. Both clauses are treated like facts; we could paraphrase the use of the comparative reference as follows: the fact that my husband calls me “wife” is better than the fact of him calling me “woman”. In other words, two expressions are compared, *wife* and *woman*, and their interrelation is formally indicated by the cohesive trigger *better*.

In example (47), it is much more difficult to identify the comparative relationship induced by the comparative adjective *better*. While the latter evidently refers to the health of N which acts as the clause’s subject and agent, the other element to which we may compare N’s current state of health eludes us. It is also missing from previous sentences. In fact, the information we seek to resolve the comparative relationship can only be retrieved from previous paragraphs. Given the absence of the third party involved in the comparative reference, we have to ask ourselves whether to include or exclude such instances of comparative reference in our study. Since these specific types of relations were quite numerous in the corpus, I have included them in the analysis although they have a rather broad cohesive scope which exceeds the regular intersentential focus generally applied. If the co-referent of a comparative reference could be identified within the entire entry, the relation was considered as being cohesive. If the referent could not be located, the relations was considered as exophoric and dismissed.

Strictly speaking, the category of the inflectional comparative marker {er} is thus slightly misrepresented in table 14 because it portrays all narrow (intraclausal), regular (interclausal) and wide (entry-wide) relations of this category. If we subtract all relations with a narrow (intraclausal) scope from the hits elicited in the entry corpus, the remaining hits for interclausal relations does not exceed 52 hits.

The third most frequent category of comparative reference in weblog entries is the lexical item *other*. Due to their semantic similitude, we may combine the comparative sub-categories of *other* and *another* to form one shared category of comparative reference; which amounts to 133 hits in all weblog entries. Accordingly, this joint category would then become the largest group of comparative reference. In addition, most instances were found to connect interclausal lexical items. This fact underscores the crucial role of the forms *other* and *another* for the construal of interclausal cohesion in weblog entries. Moreover, the syntactical construction of the comparative marker *as* + *quantifier/ adverb/ adjective* was found in all weblog entries.

Interclausal cohesive relations are exceedingly rare in this category; the cohesive scope of *as + quantifier/ adverb/ adjective* is very narrow. Example (48), for instance, shows a common finding in the corpus. Note how the cohesive pair *eyes - pie plates* is evoked by the comparative cue *as + adjective* (here: *as + big*). All three elements of this tripartite structure (cohesive pair and cue) are located within one single clause. A similar case of intraclausal comparative relations can be seen in example (49).

(48) It's all so wonderful and great! And she tries to convince us that Casey (Baby Daddy), with eyes **as big** as pie plates and a grimace that's supposed to pass for a grin, is not going to bolt as soon as the spotlight is off the trio. Wait a second.

(AA/160708/wonshpjlotc)

(49) I just want my boy to be able to enjoy life. To not have to fear something **as pervasive** as FOOD, for God's sake.

(AA/280708/tpwigapoofaa)

In contrast to examples (48) and (49), the following excerpt from a weblog entry illustrates that the category can equally set up interclausal relationships:

(50) Luckily we tend overlook, or at least forget, that Lady Liberty came from France! She's just All-American to most of us! Hope she is **as proud** of me as I am of her.

(MM/040708/hba)

The cohesive bond in (50) could be paraphrased as: “she is proud as I am proud”. It connects the pronouns *she* and *I*. They are located in adjacent clauses. In addition, they are related to each other via the comparative marker *as + adjective*. What we need to bear in mind is that a great part of the automatic hits for *as + adjective* actually produced hits which could not be characterised as comparative reference but as regular conjunctions. For instance, the potential hits *as long as* and *as soon as* clearly act as temporal conjunctions just as the construction *as well as* has to be interpreted as an instance of the additive type of conjunctive category. In all of these cases, one central fact helps to differentiate comparative from conjunctive uses of *as*

+ *adjective*: the conjunctions operate on a different scope than comparative reference. They do not merely connect noun phrases but rather conjoin entire clauses.

The conjunctive uses were consequently deleted from the findings to acquire a clearer picture of the use of comparative reference across the weblogs. This measure narrowed the general score down to 45 findings. Next, we could have factored out additionally all intraclausal comparative relations, which would reduce the overall number of hits further to less than fifteen hits in the entire corpus. It was assumed that this further reduction was not necessary and intraclausal relations were maintained in the final calculation of frequencies. This means, of course, that we need to be aware of the fact that comparative reference operates primarily on the intraclausal plane and only partially extends to the interclausal realm. This fact needs to be kept in mind when we continue to evaluate the results of comparative reference presently.

On the basis of this short analysis, I contend that while comparative reference plays a viable part in the construction of intraclausal cohesion in weblog entries, its contribution and effect on the interclausal plane of cohesion is greatly limited. The remaining sub-categories, i.e. *same*, *less*, *similar*, *such* and *equal* are equally scarce in the entry corpus. Among them, the comparative marker *same* still represents the largest category with 29 hits. We can infer the following results for the cohesive use of comparative reference in the AWC corpus A for weblog entries:

- In weblog entries, the cohesive categories of personal and demonstrative reference are far more frequent than comparative reference.
- With respect to the distribution of endophoric and exophoric reference in all weblog entries, the findings suggest that weblog entries rather exhibit features typically formed in written monologues like the genre *humanities academic prose* in Biber's study rather than spoken dialogues like conversations.
- Comparative reference is not as frequent as personal and demonstrative reference with frequencies ranging from six to twelve hits per 1000 words. As far as frequency is concerned, this finding locates comparative reference between demonstrative endophoric and demonstrative exophoric reference.

- The two weblogs which scored the most hits of all the weblogs did not show particularly high frequencies for each formal sub-category but rather had consistent frequencies in multiple formal sub-categories.
- The weblogs *Shelli's Sentiments* (ShSe) and *NonSoccerMom* (NSM) produced the highest frequency ratios of comparative reference, the weblog *All Adither* (AA) and *Back Home Again* (BHA) had comparatively low scores.
- Three out of ten main formal sub-categories of comparative reference are well represented in the entry corpus: the inflectional comparative marker {ER} as well as the comparative forms *another/other*, *more* and *as + quantifier/ adjective/ adverb*. The most salient comparative markers are the lexical items *more* and the comparative forms *another/other*.
- Some subcategories, for instance the inflectional comparative {ER} or the formal marker *as + adjective*, often induced intraclausal cohesive relations which proved that comparative reference can set up both intraclausal and interclausal relations.

In the following, these first results of personal, demonstrative and comparative reference in the AWC weblog entries will be compared with the results of reference in weblog comments. The previous findings have suggested that weblog entries are rather monologic as far as their use of reference is concerned. Next, I will investigate the monologic or dialogic dimension of reference in weblog comments.

## 6.2 Reference in Weblog Comments

It seems useful to compare the use of personal and demonstrative reference in weblog entries with the results of personal and demonstrative reference in comments. Table 15 presents such a direct comparison of eight weblog entries and eight weblog comments for these two categories. In contrast to weblog entries, there are four different types of personal and demonstrative reference in weblog comments. These are

- (a) reference within a single comment (*internal reference*)
- (b) reference between two comments (*cross-comment reference*)
- (c) reference between comment and entry (*entry-related reference*)
- (d) exophoric reference (to external parts of the weblog, shared knowledge or the discourse situation).

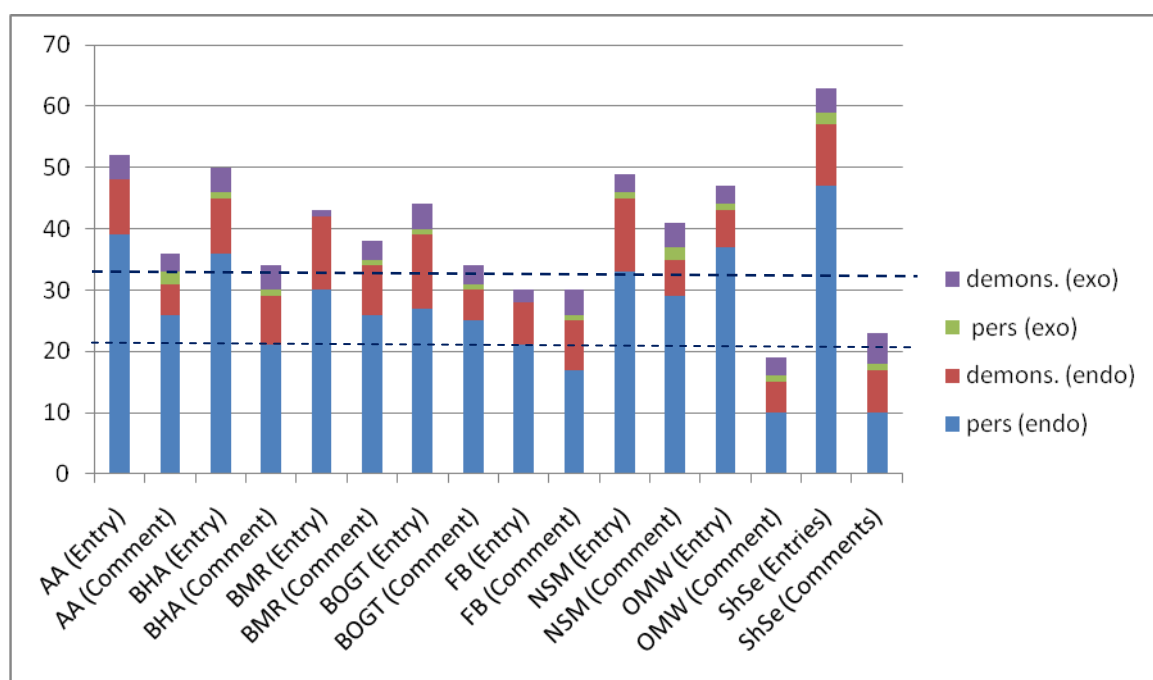


Table 15: Entries vs. Comments (Endophoric and Exophoric Reference)

Table 15 shows only relations (a) within a single comment and (b) reference between comments to account for the endophoric potential of the comment section. Relation (c) and (d) have not been included in the graph on the assumption that (c) and (d) do not actually comprise endophoric relations located inside the comment text section. We can see that on the whole, the ratio for personal reference is higher in weblog entries than in weblog comments. In most weblog comments there are around five to fifteen percent fewer items of personal reference than in corresponding weblog entries. Only two weblogs are different in this context: the weblogs *Oh my words* (OMW) and *Shelli's Sentiments* (ShSe) have higher ratings for personal reference in comments than in entries. One explanation for this exceptional finding is the fact that the comment section of the weblog *Shelli's Sentiments* features a large number of very short responses by different users. Consider the following random excerpt from this weblog's comment section:

(51) **Comment 1**

Happy Birthday, Isabelle!!!

(52) **Comment 2:**

Happy Birthday Isabelle!!! Hope you have a wonderful day and eat lots of cake and ice cream!!

(53) **Comment 3:**

Happy Birthday Isabelle! Have Grandma Shelli do a video to post here of you telling us about your birthday party.

(54) **Comment 4:**

Happy Birthday Isabelle!

(55) **Comment 5:**

Happy Birthday Isabelle - this is going to be one fantastic year for you! May all your birthday wishes come true!

(all examples: SS/050708/sgt)

In contrast to the weblog entries, the comments in the AWC are quite short. Most of the times, they are no longer than three sentences. Some of them even consist of simple clauses or even short phrases or single expressions. The brevity of the comments can be connected to the few numbers of personal references in this weblog. While the other weblogs feature quite elaborate descriptions and evaluations in their comments, the ShSe weblog mainly consists of shorter responses in the comment section. In addition, it was found that the weblog author, Shelli, frequently engaged in the comment section herself in an attempt to respond to each single comment voiced by her readers. Comments posted by weblog authors shall henceforth be called *blogger comments*. They are commonly characterised by brief acknowledgements which show only few instances of personal reference. Nonetheless, blogger comments often entail demonstrative reference. Especially the demonstratives *this* and *that* were found to be frequently used in blogger comments, referring to various referents introduced in previous user comments. Still, compared to the entry section, the amount of personal and demonstrative reference in this weblog's comment section is lower than in its entry section. This can be explained by the fragmentary nature of the comment section, which arises from the sequential synthesis of user and blogger comments. One result of this additive fragmentary style

of comment writing is a peculiar blogging style which is reminiscent of question and answer sessions in bulletin boards. At first, users respond successively to a respective entry:

Happy Birthday, Isabelle!!!	Blogarita
Happy Birthday Isabelle!!! Hope you have a wonderful day and eat lots of cake and ice cream!!	Trishk
Happy Birthday Isabelle! Have Grandma Shelli do a video to post here of you telling us about your birthday party.	Peggy
Happy Birthday Isabelle!	Avitable
Happy Birthday Isabelle - this is going to be one fantastic year for you! May all your birthday wishes come true!	Fantastagirl
Now? Now that she's five, can she finally have that cake decorating kit? Happy happy birthday to that miraculous little dolly!	Suzi
Alnternetw! Happy birthday, Isabelle!!! Five is a great year! :d	Poppy
I missed it??? Happy Birthday, Isabelle! <:-p You look like a little covergirl!	Metalmom

Table 16: User Comments in the Weblog *Shelli's Sentiments* (ShSe)

The blogger (Shelli) responds with series of blogger comments in which she reacts to all previous user comments. The number of blogger comments can be said to impose on the average length of each of these comments:

Blogarita – I am going to tell her that you and Sparky sent birthday wishes.	Shelli
TrishK – It was hard to get her to eat anything, she was so excited. She was a sweaty, happy, tired girl at the end of the day.	Shelli
Peggy – I think I may do just that.	Shelli
Avitable – Thank you, I will tell her.	Shelli
FG – I will tell her. I can hardly believe she will be going to kindergarten this year. It's terrifying for me.	Shelli
Suzi – I'm running out to get it for her in the morning. She will be here about 12:30 after her dance class. Then I am going to make a cake and help her decorate it with her new cake decorating kit. I hope they still have it at Target.	Shelli
Poppy – I know 5 is a great year. She will probably learn how to read this year. *sniff, sniff*	Shelli
MM – She looks like a little diva, but she really is too sweet to be one. She's just <i>that</i> cute. :)	Shelli

Table 17: Serial Blogger Comments in the Weblog *Shelli's Sentiments* (ShSe)

Previously, I identified two possible reasons for low frequencies for personal reference in comments, namely low word length comments and frequent blogger comments. However, both of these features fail to account for the low frequencies in the weblog *Oh my word* (OMW). For one, the latter contains long comments exceeding an average length of three sentences. More precisely, the average word length of OMW is even higher than the average word length of weblogs which showed high frequencies for personal reference. Consider the table 18. The weblog *Oh my word* (OMW) proves to have a higher average word length than the weblog *Back Home Again* (BHA) which exhibits a high frequency.

<i>BHA (high frequency of reference)</i>	<i>OMW (low frequency of reference)</i>
Average Word Length (Comments): <b>34</b>	Average Word Length (Comments): <b>47</b>

Table 18: Weblogs with High and Low Reference Frequencies (Average Word Length in Comments)

Additionally, the weblog OMW does not consist of frequent blogger comments which we reported for the weblog ShSe. There must be another reason explaining the low frequency of personal and demonstrative reference in this weblog.

One could suspect a high number of comments (per entry) to reduce the average word length of each comment which, in turn, imposes on the average frequency of personal and demonstrative reference. In order to test this hypothesis, two weblogs with high reference frequencies in comment sections were compared to the two previous weblogs which showed low frequency ratios. The results, which are illustrated in table 19, proved that we have no empirical basis for the above claim. The amount of comments per entry showed varied considerably from one weblog to the next. Hence, comment numbers do not correspond to reference frequencies.

<i>AA (high frequ.)</i>	<i>BHA (high frequ.)</i>	<i>ShSe (low frequ.)</i>	<i>OMW (low frequ.)</i>
25	9	9	15

Table 19: Average Number of Comments per Entry in Weblogs

More reasons for the low reference frequency in the weblog may be found on a qualitative plane of analysis. Let us therefore compare two more complex comments



drawn from two different weblogs. One is an excerpt from the weblog BHA, which shows a high frequency of reference relations, the other one is taken from the weblog OMW, which exhibits a low frequency ratio:

(56) I'm glad for your review of the Tekamah Motel, Jerry. We have never stayed **there** (never had to, Mom and Dad had a house in town back then) and when Dad moved to Okland (did you go **there**, it's a Swedish town?) we stayed at one of **their two motels**. At any rate, I hear **it** has been updated. When I was doing road work for a summer job, a lot of the road workers were staying up there at the Tekamah Motel. We built the paved highway halfway to Oakland. I'm glad your Dad got a lift, it is too hot for **him** to be walking that far right now.  
(BHA/160708/b08gt)

(57) It's not necessary to be funny all the time - As those before me have said, very, very well written. I share your feelings about the direction the world may be headed; at times it can easily become overwhelming, even for the most upbeat of us. It pains me to no end to see my (and other sensible Americans') hard-earned taxes quickly spent on such horrors. The shame felt at simply being American these days is truly powerful sometimes, as well. I like to think that things in fact will change, and soon; I have hope that the international community is simply waiting for Bush to vacate the office, at which point **they'll** be poised to give us perhaps one more chance.  
(OMW/130708/som)

Note that the first comment includes multiple instances of interclausal personal and demonstrative reference. It deals with the user's past experience, which is formally underscored by the frequent use of past tense verbs (e.g. *had*, *moved*, *built*, *stayed*, *got*) perfect aspect verbs (e.g. *was doing*, *were staying*) as well as the consistent use of first person pronouns (singular and plural). The comment thus builds up a distinctive narrative profile. By contrast, the second comment only comprises one instance of interclausal reference. It does not share the same narrative features as the first comment. Past tense verbs are missing. Instead, it mostly features the use of present tense verbs, which mirror the author's current view on the state of affairs.

This is not to suggest that the weblog does not comprise any narrative comments but rather, that these are underrepresented in this weblog's comments. The absence of a narrative profile in the second comment coincides with its low average frequency of personal and demonstrative reference. I suppose that there is a quantitative correspondence between the frequency of formal parameters, which characterise narrative discourse, and the frequency of personal and demonstrative reference. This explains the low frequency of personal and demonstrative reference in the weblog *OMW* which lacks the generic narrative profile we have found in the comment sections of other weblogs. In other words, the discourse in the comment section of the weblog *OMW* is assertive and argumentative rather than narrative.

The weblog *Five Blondes* (FB) is the only weblog in the corpus with similar reference frequencies in both entries and comments. Remember that this weblog is written collaboratively by four sisters. A closer look at the weblog's comments reveals that all four authors of the weblog are the main commentators to each other's entries. If blog entries and comments are basically written and composed by the same authors, it is not surprising that the general frequencies for reference in the entry and comment sections converge.

Let us now take a look at the frequencies for personal and demonstrative reference between weblog comments and entries. The referential rapport between entries and comments will shed some more light on the dialogical status of the personal weblogs. Table 20 illustrates the four types of referential relations grouped according to their standardized frequencies in eight different weblog comment sections:<sup>73</sup>

- internal reference (anaphoric and cataphoric reference within comments)
- entry reference (reference between entries and comments)
- exophoric reference (reference to external weblog elements or to shared knowledge of interlocutors)

We can see that internal reference (anaphoric and cataphoric reference) is the most frequent category in weblog comments. Next are relations between comments and entries (entry), followed by exophoric relations and cross-comment relations. The majority of weblogs exhibit similar ratios for internal reference except the two weblogs

---

<sup>73</sup> The analysis of four different reference relations proved to be particularly cumbersome and time-consuming. As a consequence, only eight out of ten weblogs were selected for this demanding stage of the referential analysis.

OMW and ShSe for reasons discussed previously. The frequencies for exophoric relations are comparatively low but stable across the weblogs. There are only few cross-comment relations throughout the weblogs' comment sections. Most cross-comment relations could be elicited in the weblog *Beetle's Memories and Ravings* (BMR), but even these relations appear minor when compared to the other reference relations. The low overall frequencies for cross-comment relations indicate a minor degree of interaction between weblog comments. As we will see, complex interactive exchanges, which are well documented for internet chats, for instance, do not emerge in weblog comments.

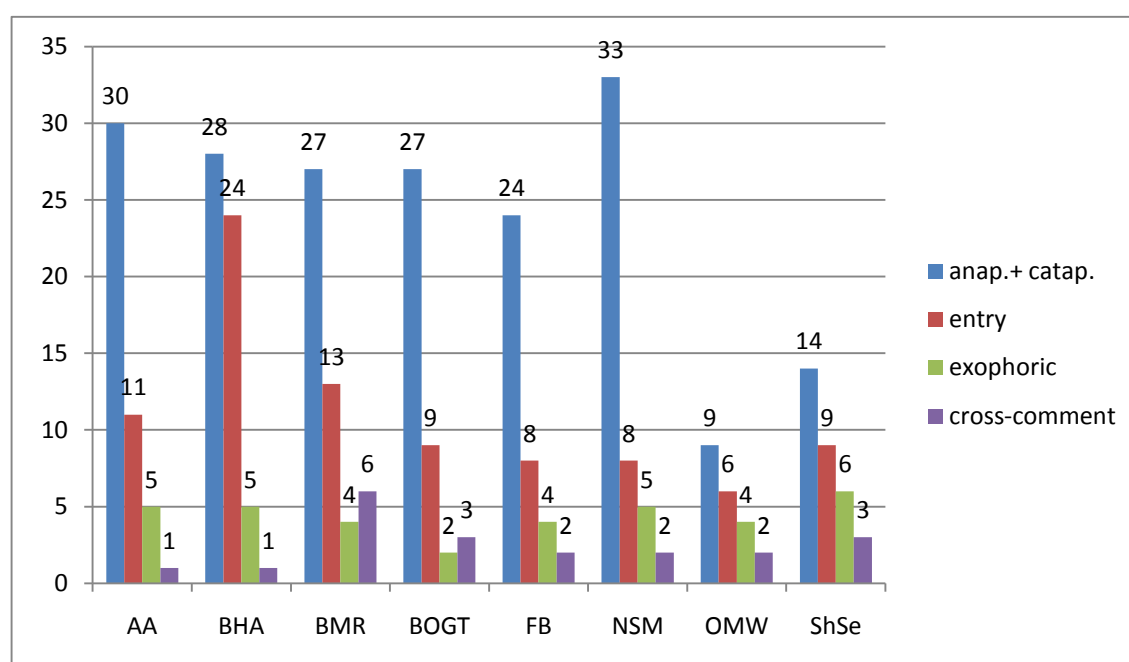


Table 20: Personal and Demonstrative Reference in Weblog Comments

Similar to the results of weblog entries, exophoric reference relations were time-related and therefore connected to time stamps or date indicators retrievable from the comment itself or from the entry to which it refers.

Finally, let me evaluate the results for reference relations between comments and entries. These relations are particularly interesting for this study, as they indicate the level of cohesive rapport between weblog entries and comments. The mean score for this category of relations is 9 h/w for all weblogs except for the weblog *Back Home Again* (BHA). It shows unusually high frequencies in this category. How can this difference be accounted for? In contrast to all of the other weblogs in the AWC, the author of the weblog BHA regularly posts entries about his favourite music

albums. These entries are habitually staged in a particular fashion: first a picture and description of a music album is presented. Then, one song is chosen from the album and featured prominently in the entry. The entry usually ends with an embedded video or audio file of the respective feature song for users to view or listen to. These relatively short “music entries” thus mainly centre on one prominent song of a music group which is discussed in detail. The following figure illustrates the text design of a regular music entry.

#### Jane - Jefferson Starship



Title: [Jane](#)  
Artist: [Jefferson Starship](#)  
Composers: [David Freiberg](#) / [Jim McPherson](#) / [Paul Kantner](#) / [Craig Chaquico](#)  
Label: [Grunut 11750](#)  
Year Released: [1979/1980](#)  
Highest Chart Position: [14](#)

After a two week hiatus traveling around the country for blogmeets, [Friday Sock Hop](#) returns....

No particular reason for this song. I just always liked it. I like [Mickey Thomas](#) the best of all the lead singers in the Jefferson Airplane / Jefferson Starship / Starship bands. He also sang lead on the Elvin Bishop classic *Fooled Around and Fell in Love* in 1976. What a great voice.

I've featured Starship before on Friday Sock Hop. They were voted as the #1 song of the 50 Worst Songs Ever with their hit [We Built This City](#). I have the 45. That was a #1 hit from 1985. I think many others bought it too.

I've included the video for Sara too. Look for [Rebecca de Mornay](#) in that video. Enjoy!

Enjoy!

[Play \(5.9 MB\)](#)

[Watch Video \(Jane\)](#)

[Watch Video \(Sara\)](#)

---

Figure 30: A Prototypical Music Entry  
(Featuring Album Picture, Song Description and Audio/Video File)

The majority of comment-entry relations in this weblog could be elicited in comments which responded to these music entries. There are two main reasons why commentators may have been prone to use entry-related reference in this specific context. On the one hand, the music entries are quite short. Comments are automatically posted beneath the entries, so that short entries enable a quick identification of cohesive referents between the respective comment and its entry. On the other hand, most entry-related reference referred to the song, which is made available via audio or video files at the end of the music entry.

You remember the Righteous Bros? Yeah.

The one with the high voice?. Yeah.

Well, he died somewhere in this room. Probably on that bed or in that bathroom.

Here's a video of Bobby singing his biggest hit *Unchained Melody*. If there's a [Rock and Roll Heaven](#), well you know they've got a hell of a band. RIP, Bobby!



(enlarged segment)

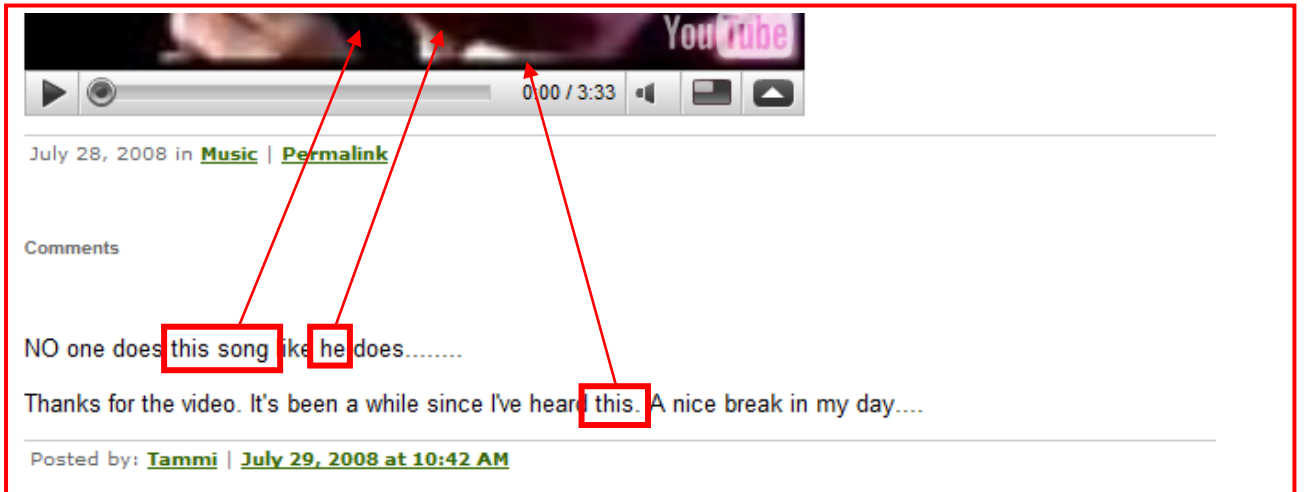


Figure 31: Entry-Related Reference in Comments Following a Music Entry

The song is therefore not only the most salient topic addressed in the music entry, it also functions as its final referent. It is quite natural for commentators to address this referent in their posts, not least because it is the most proximate and therefore most likely candidate for reference resolution. One particular music entry in the BHA weblog closes with an embedded (youtube) video. Its picture shows the large counterfeit of the singer-songwriter Bobby Hatfield alongside a verbal caption of his song "Unchained Melody". The visual salience of this video triggers a series of comments which refer to it via demonstrative reference.

On a formal plane, comment-entry relations are dominated by demonstratives rather than personal reference. The relations usually refer to large pictures or videos. Therefore, we can contend that visual stimuli frequently induce entry-related

references. In fact, if we subtract all visual referents from the list of comment-based entry relations, the number of hits would decrease considerably, showing about one third less hits than before.

Generally speaking, we can therefore explain the unusually high frequency of entry-related comment relations by recourse to the brevity of BHA's music entries, their topical simplicity, their use of visual stimuli as well as the spatial proximity between comment-based anaphora and entry-based antecedent. Other weblogs either used longer or thematically complex entries which complicated anaphora resolution across the text units or they did not emphasise one topical element of the discourse as saliently as the music entries in the weblog *Back Home Again* (BHA). Most commonly, however, we can say that entry-related reference relations are not very frequent in personal weblog comments. In comparison to internal reference, entry-related references seem to play a minor role in most weblog comments. I suggest that this is a first indicator of limited interaction between weblog entries and comments. Although there appears to be some cohesive rapport between weblog comments and entries, cohesive interaction is rather limited and one-sided. Whereas examples of comments referring to entries could be found repeatedly in the corpus, I could not identify personal or demonstrative cues in entries actively referring to weblog comments.

Let us now focus our attention once more on the use of comparative reference in the AWC weblog comments. The tables 21 and 22 illustrate the general frequencies for comparative reference in eight AWC weblogs across entry and comment sections. The most remarkable difference between weblogs and comments concerns the frequencies for the weblog *All Adither* (AA). The frequency of comparative reference has greatly increased from entry to comment sections. While it exhibited one of the lowest frequencies for comparative reference in the entry sections, it has the highest frequency of all weblog comment sections. Apart from this weblog, all other weblogs only exhibit slight changes in their frequencies. On the one hand, weblogs which featured high frequencies for comparative reference in their entry section have somewhat lower frequencies in their comment section. On the other hand, weblogs which had lower frequencies in their entry section now have higher frequencies in their comment section. On the whole, the frequencies for comparative reference across the weblog comments seem to be rather consistent, ranging between eight and nine hits per 1000 words. Again, we need to bear in mind

that these numbers include intraclausal relations and therefore may convey a somewhat distorted picture given this study primarily focuses on interclausal cohesive relations. Therefore, a focus on interclausal comparative relations would certainly have produced smaller frequencies than the ones presented in the following tables. Still, I believe that the basic implications voiced before are likely to be the same in both cases.

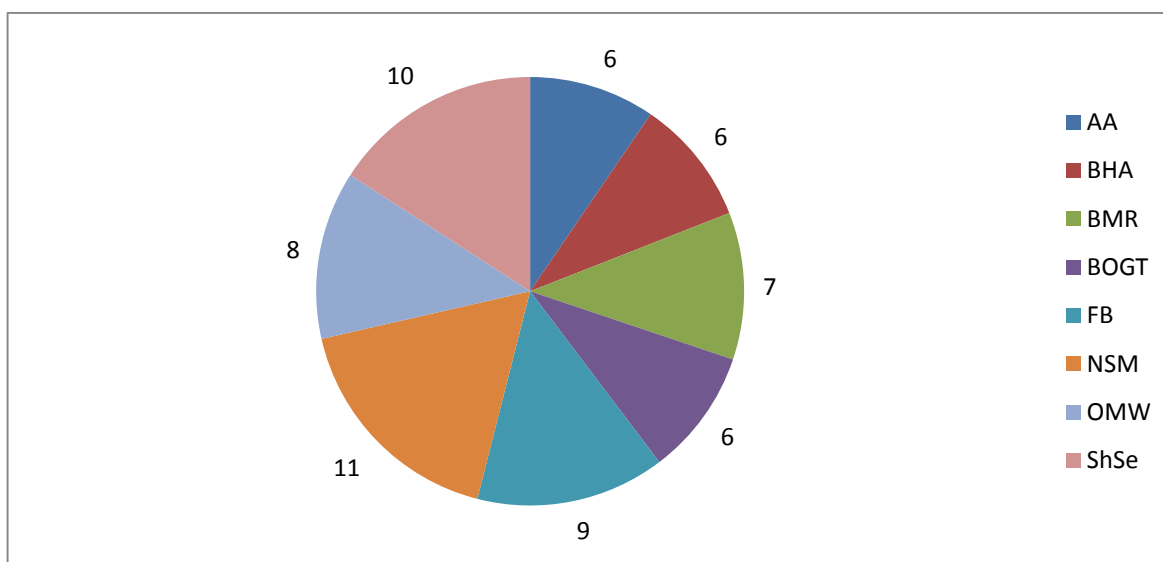


Table 21. Comparative Reference in Entries

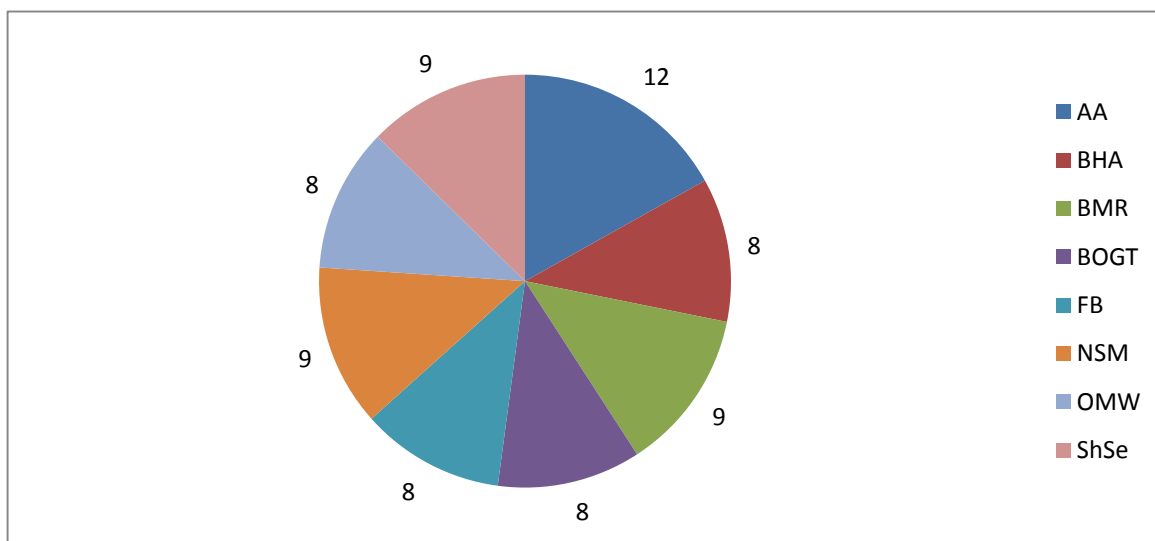


Table 22: Comparative Reference in Comments

I shall now proceed to compare the frequencies for comparative reference between weblog entries and weblog comments. Table 23 reveals the major changes between weblog entries and comments in this context. Apparently, the cohesive profile of the collaborative weblog *Five Blondes* remains constant in both entry and comment sections. This may be another quantitative indicator for the fact that entries and

comments in this weblog are basically written by the same group of authors. Both entry and comment profiles of this weblog are very much alike. Similar frequencies between weblog entries and comments may thus be an indication of relatively self-contained weblogs whose authors appear as the only (or primary) commentators.

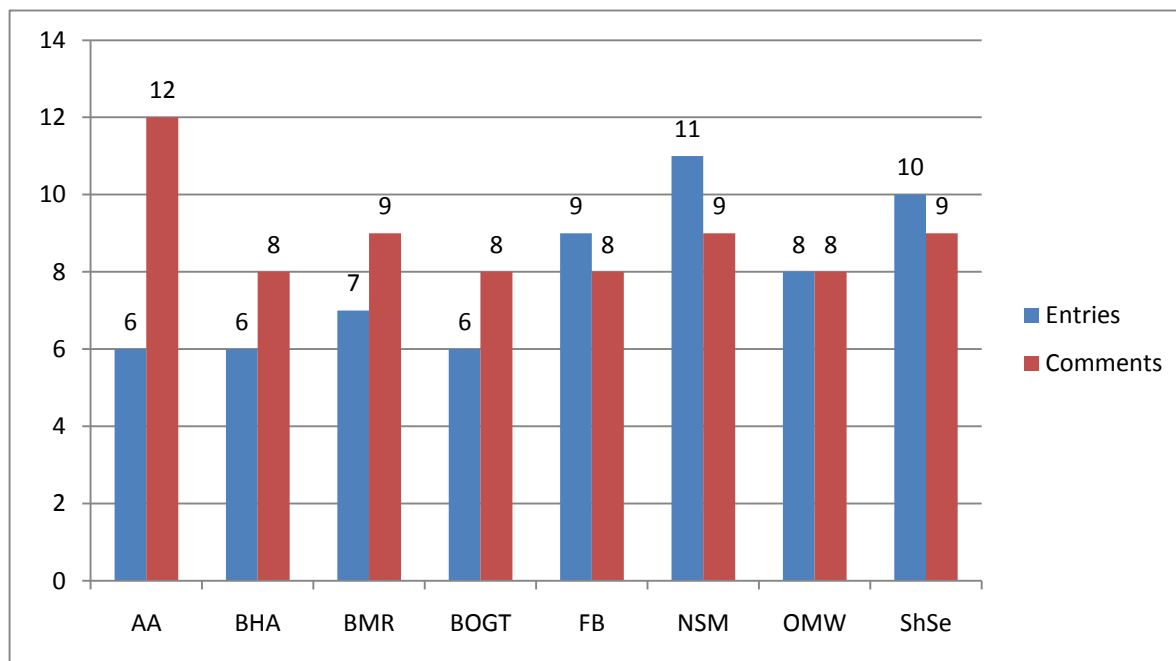


Table 23: Comparative Reference in Weblog Entries and Comments

Altogether, comparative reference in weblogs slightly increased in comments. It is conceivable that users frequently apply comparisons in comment sections to relate what they have read in entries to their own personal experience. The only two exceptions are the weblogs *Five Blondes* (FB) and *NonSoccerMom* (NSM), which have more hits for entries than for comments. However, the difference between comments and entries is minor and insubstantial.



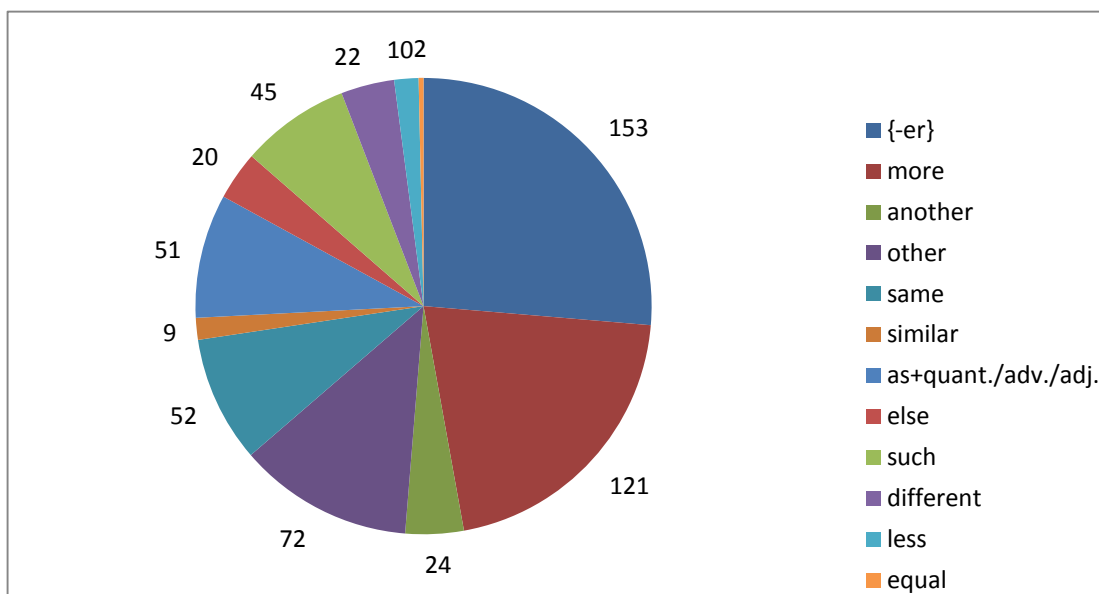


Table 24: Comparative Reference in Comments (Categories)

I shall now consider the individual categories involved in comparative reference in comments which are illustrated in table 24. It depicts the total amount of hits for each comparative category in all weblog comments. The values have not been normalized yet and can thus not be compared to the frequencies for the weblog entries. The table simply provides a broad overview of the amount of hits for each category across all comment sections. Again, similar to the entry sections, three categories are most frequent: the inflectional comparative suffix {-ER}, the lexical marker *more* as well as the two similar categories of *other* and *another*. In addition, the categories *same*, *as + quantifier/ adverb/ adjective* and *such* represent the second largest categories in the data. The remaining categories only assume a minor position compared to these two frequent groups.

We can compare the normalized frequencies for each category in the entry and comment section. The following tables show that the frequencies for each category are very low compared to other cohesive categories, such as reference (or substitution and conjunction as we shall see presently):

### Weblog Entries

<i>Categories</i>	<i>Frequency</i>
{-ER}	2
more	2
another	1
other	2
same	1
similar	0
as+quant./adv./adj.	1
else	0
such	0
different	1
less	1
equal	0

### Weblog Comments

<i>Categories</i>	<i>Frequency</i>
{-ER}	2
more	2
another	0
other	1
same	1
similar	0
as+quant./adv./adj.	1
else	0
such	1
different	0
less	0
equal	0

Tables 25 and 26: Normalized Frequencies for the Sub-Categories of Comparative Reference

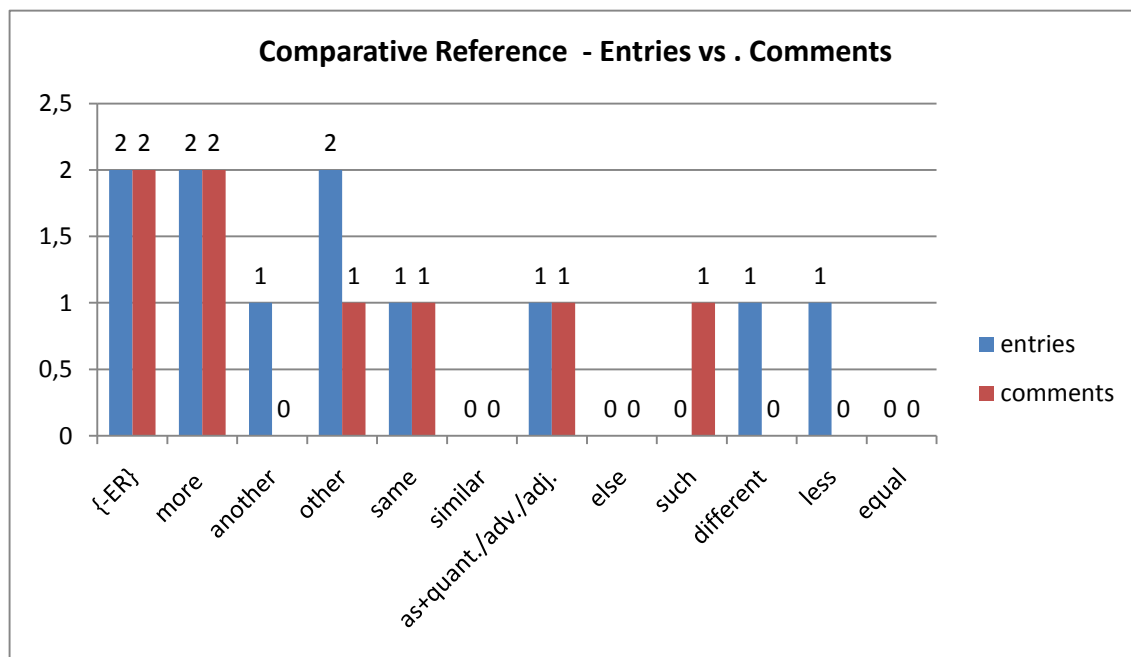


Table 27: Entries vs. Comments (Comparative Reference, h/w)

Evidently, the weblog entries include a variety of different categories with at least one or two hits per 1000 words. The comment sections consist of a more limited set of frequent categories. Some categories with insubstantial numbers are not depicted in the figure. Neither the weblog entries nor the comments show comparative

categories which exceed three hits per 1000 words. This finding underscores the previous hypothesis that comparative reference is much less frequent than personal or demonstrative reference in personal weblogs. In addition, the results affirm that the most frequent categories across weblog entries and comments are the inflectional suffix {-ER}, the comparative marker *more* as well as the categories *other/another*. There are some minor differences between the frequencies for entry and comment sections. For instance, the category *another* shows insignificant hits in the comment section which contrasts with its fairly consistent use in the entries. Its semantic sibling, *other*, equally drops its frequency ratio from two hits to only one hit per 1000 words. There is a general decrease in the use of both of these comparative markers from weblog entries to comments. It may be partially compensated by a slight increase in numbers of the category *such* in the comment section. The categories *less* and *different*, which appeared in minor numbers in the entries, hardly occur at all in comments. Generally speaking, these minor frequency shifts in weblog entries and comments are not suggestive of any crucial difference between the use of comparative reference in entries and comments. Categorical frequencies remain roughly the same with only few negligible differences.

I shall now review the central results of the evaluation of reference in the AWC weblog comments:

- In weblog comments, there are four possible types of reference, i.e. internal reference, cross-comment reference, entry-related reference and exophoric reference. While internal and exophoric reference is comparable to the endophoric and exophoric reference in entry sections, cross-comment and entry-related reference construe cohesion in a collaborative manner.
- Entry-related personal and demonstrative reference is highly constrained in most weblogs. This finding suggests a largely monologic profile of weblog comments. The cohesive interaction with corresponding entry sections is generally low, with the exception of the weblog *Back Home Again*. Its high ratio of entry-related comments can be explained by its particular text design of the “music entries” introduced previously. The latter seem to encourage anaphoric entry-related reference more than the regular entries of the other weblogs.
- Surprisingly, a considerable amount of entry-related references was found to refer to visual representations (pictures or videos) displayed in entries. These

semiotic modes seem to trigger demonstrative entry-related reference more than verbal referents. This suggests that weblog pictures can play a constitutive part in the interactive maintenance of cohesion across weblog entries and comments.

- Cross-comment reference relations between different comments are less frequent than entry-related reference. The data suggests that interaction between comments is mainly limited to blogger comments responding to user comments. Firstly, most bloggers do not engage in their comment sections very often and, secondly, they do not always make use of cross-comment reference if they do. Within comment sections, the interaction is therefore one-sided, leading from the blogger comment to the user comment. No user responses to the blogger comments (*response-responses*) could be identified in the entire corpus.
- Frequencies for exophoric reference remain roughly constant between entries and comments. As in entries, most exophoric references are time deictic references to time stamps or date indicators external to entry and comment text units.
- The collaborative weblog proved to have identical weblog authors and commentators. Therefore, the reference profile for both entry and comment section looked very similar. There is a possibility that authors of collaborative weblogs are generally prone to engage in each other's comment sections, which makes for a convergence of frequencies for both entry and comment sections in these weblogs. However, the data of this study was too limited to allow for such a conclusion.
- In weblog entries and comments comparative reference proved to be scarce. Neither entry nor comment section exhibited comparative categories, which exceeded two hits per 1000 words. As such, comparative reference is much less represented than personal and demonstrative reference in the corpus. The use of comparative reference in comment sections does not deviate substantially from the use in entry sections. On a general plane, however, comparative reference contributes to the construction of intra- and interclausal cohesion. The most pervasive categories of comparative cohesion both in the entry and comment sections of the weblogs are the inflectional suffix {-ER} and the periphrastic marker {more}, followed by the two similar categories of

*other* and *another*. Less frequent categories were *similar*, *equal*, *less*, *such* and *different*.

I shall now close this chapter by evaluating the quantitative and qualitative distribution as well as the use of substitution, ellipsis and conjunction in both entry and comment sections of the AWC weblogs.

### **6.3 Substitution in Weblog Entries and Comments**

It seems useful to begin the analysis of substitution in weblogs with an overview on the distribution of substitution in entries and comments (cf. table 28). The average frequency of substitution across the different weblogs in the AWC is comparatively low in contrast to the high frequencies for personal and demonstrative reference (approx. 10 - 40 h/w). The frequencies for substitution do not exceed four hits per 1000 words. The average frequency of substitution is therefore very similar to the distribution of comparative reference. The category is hardly more frequent than comparative reference in entry or comment sections.

There are no substantial frequency changes between entry and comment sections. The weblogs *All Adither* (AA), *Beetle's Memories and Ravings* (BMR) and *NonSoccerMom* (NSM) have most hits for substitution in the entry section. The other weblogs have an average frequency of half as many hits per 1000 words (2 h/w). The comments have a slightly larger average frequency than the entries by 2,5 hits per 1000 words. There are some differences with respect to individual weblogs. The weblog *All Adither* (AA), for instance, has twice as many hits for substitution in its entry section than in its comment section (4:2). Similarly, the weblog *Oh my word* (OMW) presents a differential of 2:1 between substitution hits in entry and comment section. In contrast, other weblogs have higher frequencies for substitution in comments than in entries. The weblogs *Back Home Again* (BHA), *Five Blondes* (FB) exhibit slightly more hits in the comments than in entries. These differences are insubstantial. Rather, the amount of substitution used in entries and comment sections is consistently low. A low frequency of substitution is indicative of a monologic rather than a dialogic setting.

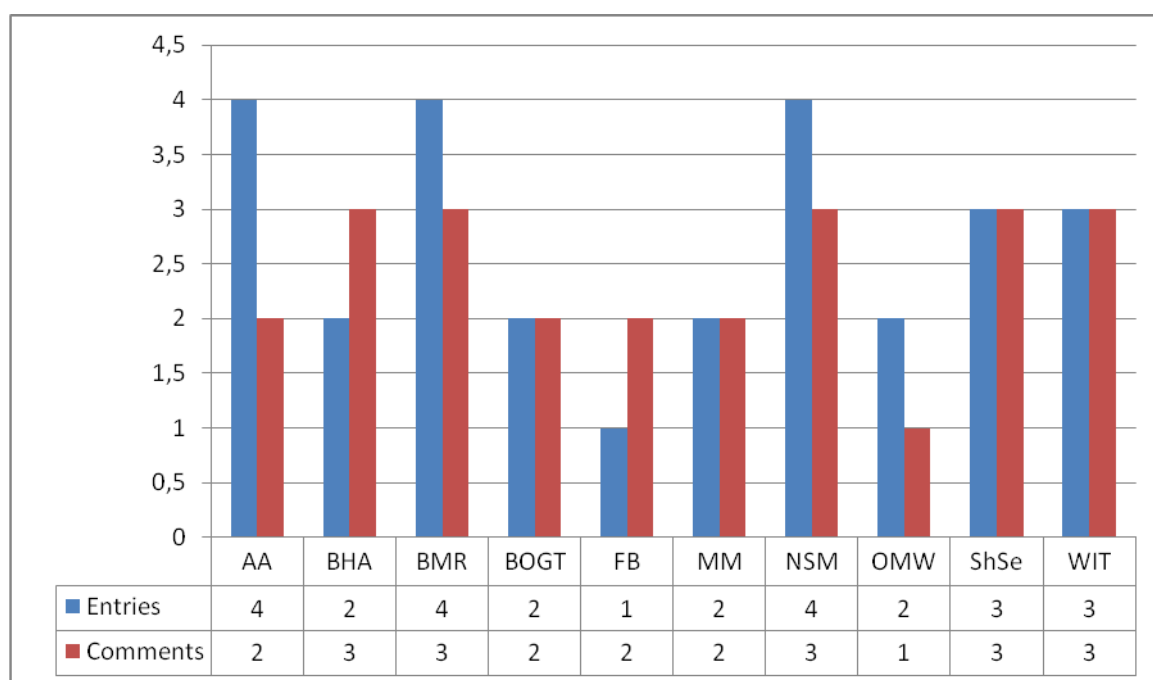


Table 28: Substitution in Weblog Entries and Comments

For instance, Brown & Yule (1983:15) inform us that “spoken [conversational] language contains many incomplete sentences, often simply sequences of phrases”.<sup>74</sup> We may take this “condensed way of speaking” as a result of the lack of planning time in unregulated oral discourse settings, such as two-party and three-party conversations. Substitutions are cohesive means which help to condense ongoing discourse by achieving the following three important communicative goals:

- (a) to extend the social rapport between interlocutors (by appropriating the speech of others by substitution), indicating shared knowledge
- (b) to reduce the amount of information which needs to be processed and interpreted by the interlocutors
- (c) to reduce the amount of planning and execution of a particular speech act.

In written monologic discourse, the amount of planning involved in contributions is much greater than in dialogic conversational settings due to the asynchronous, paper-bound mode of production and reception. We may now consider the types of substitution and their overall distribution in entry and comment sections (cf. table 29 and 30).

<sup>74</sup> For prototypical ways to distinguish spoken and written language in previous linguistic research, compare also Stubbs (1998:34)

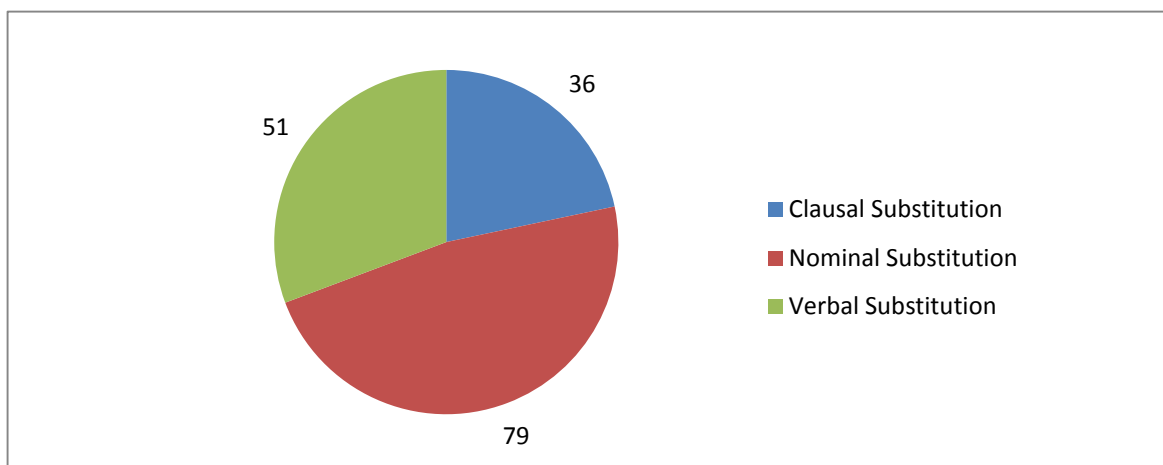


Table 29: Substitution in Entries (Categories)

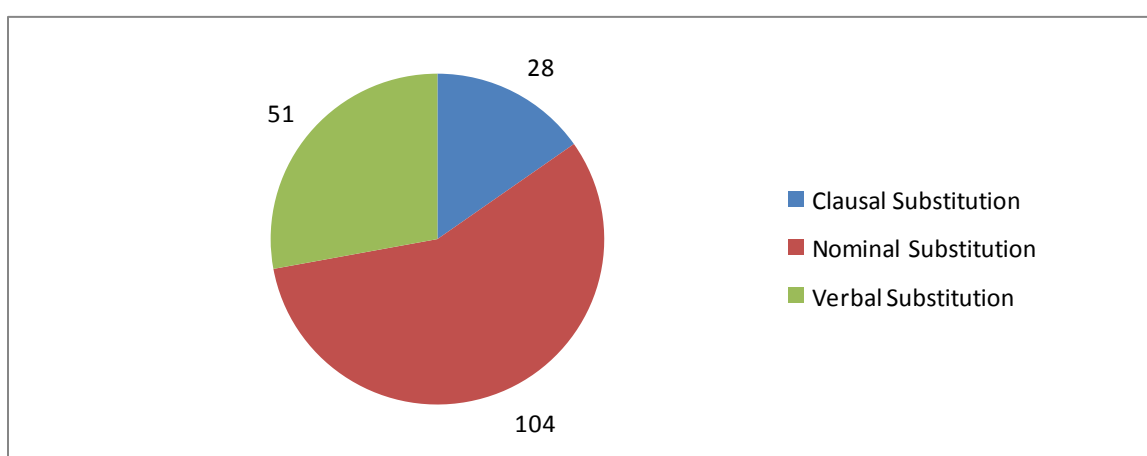


Table 30: Substitution in Comments (Categories)

The distribution of the categories of substitution in entry and comment sections is similar. Nominal substitution (especially the category *one/ones*) is the most common type of substitution in the weblogs. The number of hits in the comments is slightly higher than in the entries. The frequencies for verbal and clausal substitution show similar numbers in the entry and comment sections. Verbal cohesion is the second largest category. One quarter of all substitution relations are verbal in kind. The smallest category is clausal substitution. The frequency of clausal substitution is insignificantly higher in entry than in comment sections. These findings indicate that nominal and verbal substitutions are the two most frequent types of substitution in the AWC. The most frequent subtypes in these two classes are *one/ones* for nominal substitution and the past tense word form of the lexeme DO: *did*. Most substitution relations were interclausal; no intraclausal relations could be identified in the corpus.

Most substitution relations were intratextual, only two instances were found to be intertextual, i.e. substitutions referring to contributions of other authors/users:

- (58) Daisy: Hahahaha! I do not know why, but I really, really enjoyed hearing your Voki [speaking avatar]. Maybe it's the accent. Or the music.

Babs: I'm hoping you will make **one** also Daisy.  
(BMNR/140708/mvp)

- (59) Entry: [...] And you (and That Guy) will be happy to know that I couldn't have another cigarette if somebody paid me to smoke. There's some link in my brain between smoking and puking now, and so, I can't even look at cigarettes. You would think the same thing would happen with Leffe beer and ham and cheese or pâté sandwiches. No such luck.

Comment: OMFG! I really feel for you on this **one**. I've never been able that even think that I could handle puffing a fag EVER! I just know I'd turn green and blow chunks or shit myself!  
(OMW/050708/htqs)

The following two charts 31 and 32 reveal the internal variation between the types of substitution used in entry and comment sections across the different weblogs. Five weblogs show the same categorical variation in entries as in comment, i.e. *Beetle's Memories and Ravings* (BMR), *Five Blondes* (FB), *NonSoccerMom* (NSM), *OhmyWord* (OMW) and *Shelli's Sentiments* (ShSe). All of these weblogs exhibit nominal and verbal substitution in both entry and comments section. Some of them include small frequencies for clausal ellipsis. The remaining five weblogs show differences between the entry and comment profiles of substitution types, although these differences are not substantial in numbers. These are *All Adither* (AA), *Blog of a Good Time* (BOGT) and *What ever I think* (WIT). In two of these three weblogs (AA and BOGT), verbal substitution recedes in favour of nominal substitution. Only in the weblog WIT does verbal substitution increase and nominal substitution decrease in numbers. This latter tendency seems to be exceptional. In most weblogs nominal



substitution either increases from entry to comment sections or remains stable. The weblog *Back Home Again* (BHA) is the only weblog which shows an increase of categorical variation from entries to comments; this is, again, a rather exceptional finding. Furthermore, the weblog *Mushy's Moochings* (MM) does not display such an increase, rather its frequencies between entries and comments are identical. The difference between entries and comments in this weblog is thus not quantitative but qualitative.

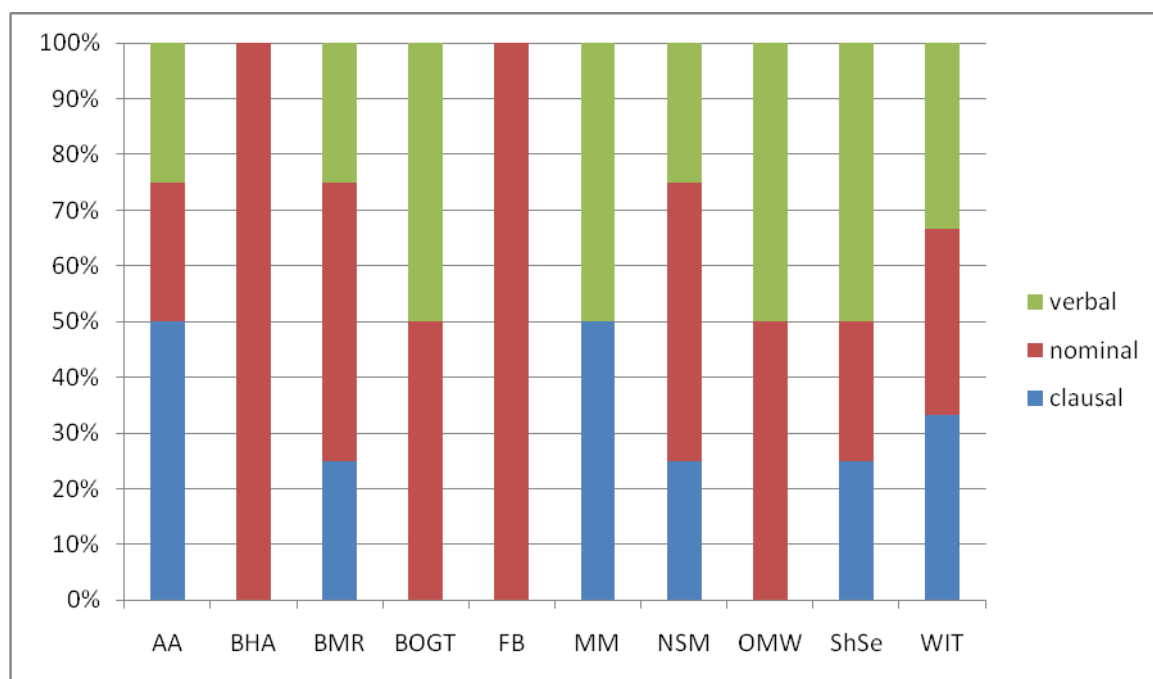


Table 31: The Distribution of Substitution Sub-Categories Across Weblog Entries

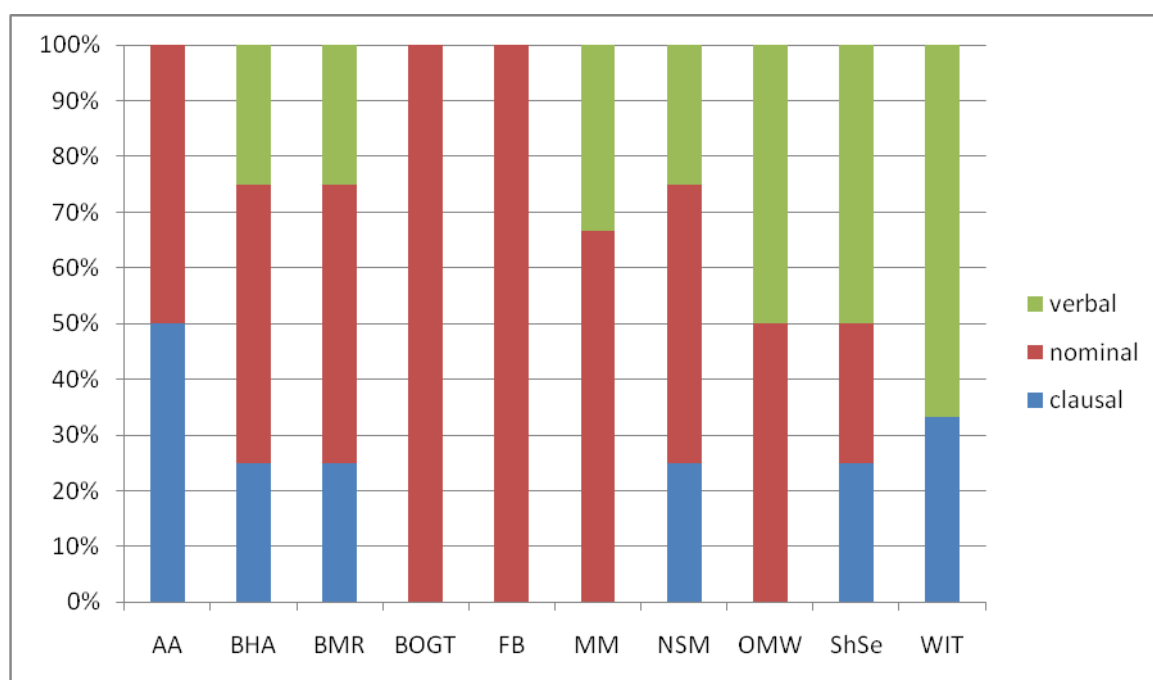


Table 32: The Distribution of Substitution Sub-Categories Across Weblog Comments

The causal substitution in the entry section of the weblog *MM* is replaced by nominal substitution in the comments. In addition, the frequency of verbal cohesion slightly decreases. This finding sustains the hypothesis that nominal substitution represents the most salient category of all three types. It is particularly frequent in comment sections.

#### 6.4 Conjunction in Weblog Entries and Comments

In this section, I will now consider the distribution of the cohesive category of *conjunction* in the AWC. Table 33 provides a rough outlook on the frequencies for *conjunction* in and across weblog entries and comments. It is evident that *conjunction* is used more frequently in weblog comments than in entries. In some weblogs, this difference is quite subtle, e.g. in *Mushy's Moochings* (MM) and *Ohmyword* (OMW). In other weblogs, the contrast is stark, e.g. *Back Home Again* (BHA), *Whatever I think* (WIT).

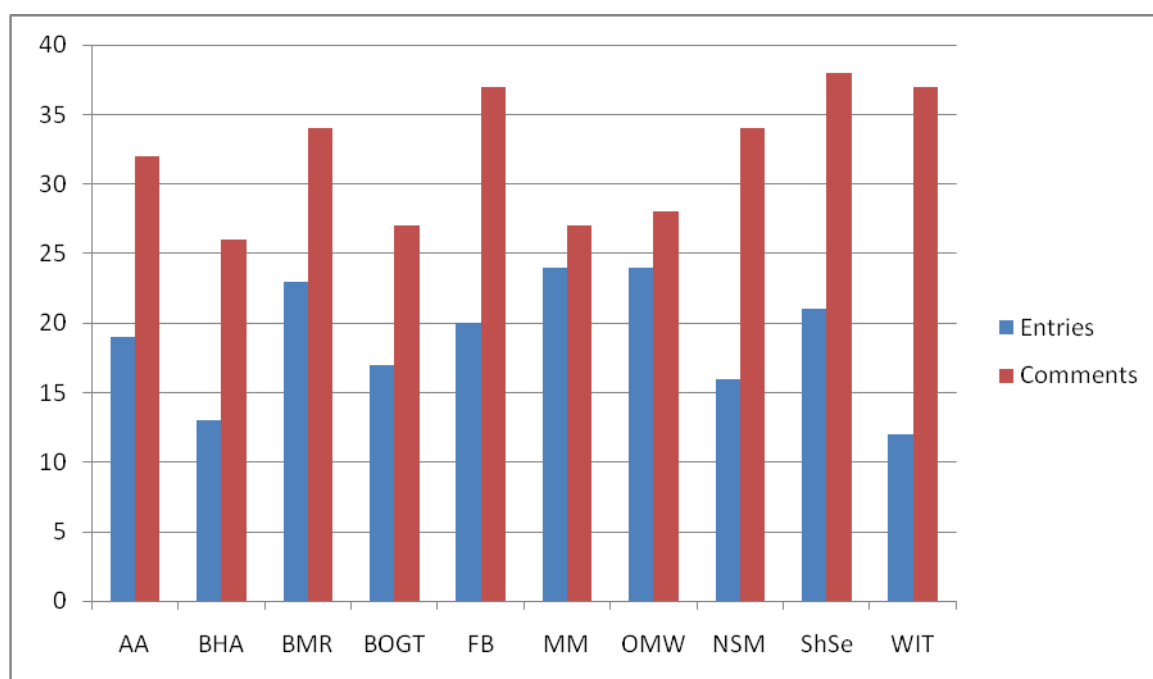


Table 33. Conjunction in Weblog Entries and Comments

Let us take a closer look at the types of conjunction and their distribution across weblog entries and comments. The first in a row of conjunctive categories is *additive conjunction* (cf. table 34). Both the entry and the comment sections have comparatively high frequencies compared to the three other main conjunctive categories (*adversative*, *causal* and *temporal conjunction*) in the AWC. Additive

conjunction represents the most frequent type of conjunction in both weblog entries and comments. With the exception of one weblog (BHA), the use of additive conjunction is more frequent in the entry than in comment section. This finding seems to be directly related to the average word length of each entry or comment sections. As the entries are usually longer than the comments, more coordinating conjunctions are used to conjoin compound and complex clauses. The only weblog which shows higher frequencies for additive conjunction in its comments than in its entries is the weblog *Whatever I think* (WIT). The average word length of comments in this weblog is larger than in other weblogs. This confirms my previous assumption that a high average word length of text units correlates with the frequency of additive conjunctions. The high average word length of the comment section may, in turn, be explained by the high frequency of relatively elaborate blogger comments in this weblog. This regular usage of blogger comments in the weblog WIT is similar to the high frequency of blogger comments in the weblog *Shelli's Sentiments* (ShSe). If we take a look at the frequency of additive conjunction in this weblog, we realise that the weblog ShSe actually has the highest frequency of this category across all comment sections; a finding which further sustains the previous argumentation.

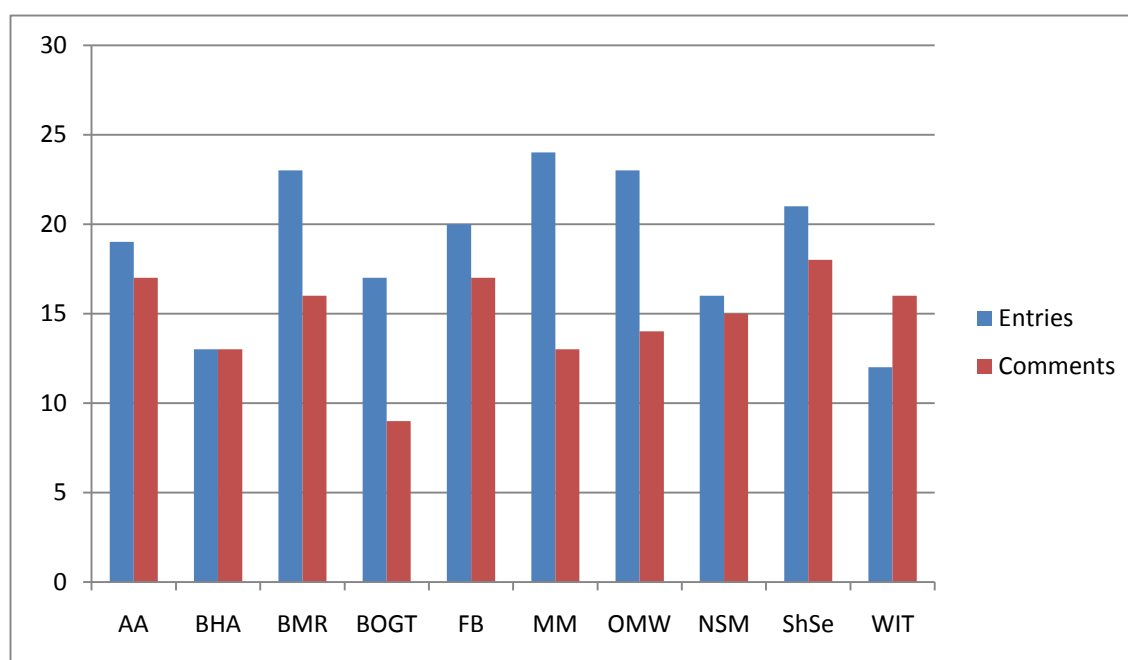


Table.34: Additive Conjunction in Weblog Entries and Comments

The second most frequent type of conjunction in weblog entries and comments is *adversative conjunction*. It is the type of conjunction prominently used to express an opposition of stance or contrastive meaning between propositions. Smith & Frawley

(1983:366) have demonstrated that adversative conjunctions are more frequently used in text genres which deal with “the expression of events in time” rather than in text types which are more argumentative in nature, e.g academic articles. They even claim that high frequencies for additive and adversative conjunctions are characteristic of narrative genres like fiction (cf. table 35).

<i>Rank/Genre</i>	<b>Fiction</b>	<b>Journalism</b>	<b>Religion</b>	<b>Science</b>
<b>1.</b>	Adversative	Adversative	Additive	Additive
<b>2.</b>	Additive	Temporal	Adversative	Adversative
<b>3.</b>	Causal	Additive	Causal	Causal
<b>4.</b>	Temporal	Causal	Temporal	Temporal

Table 35; Cohesive Conjunction Types  
(Frequency Ranking According to Smith & Frawley 1983:365)

In table 35, we can see that in Smith & Frawley’s study, the text genres fiction and journalism showed most hits for additive conjunctions followed by religion. Scientific texts showed the least hits of all four text genres.

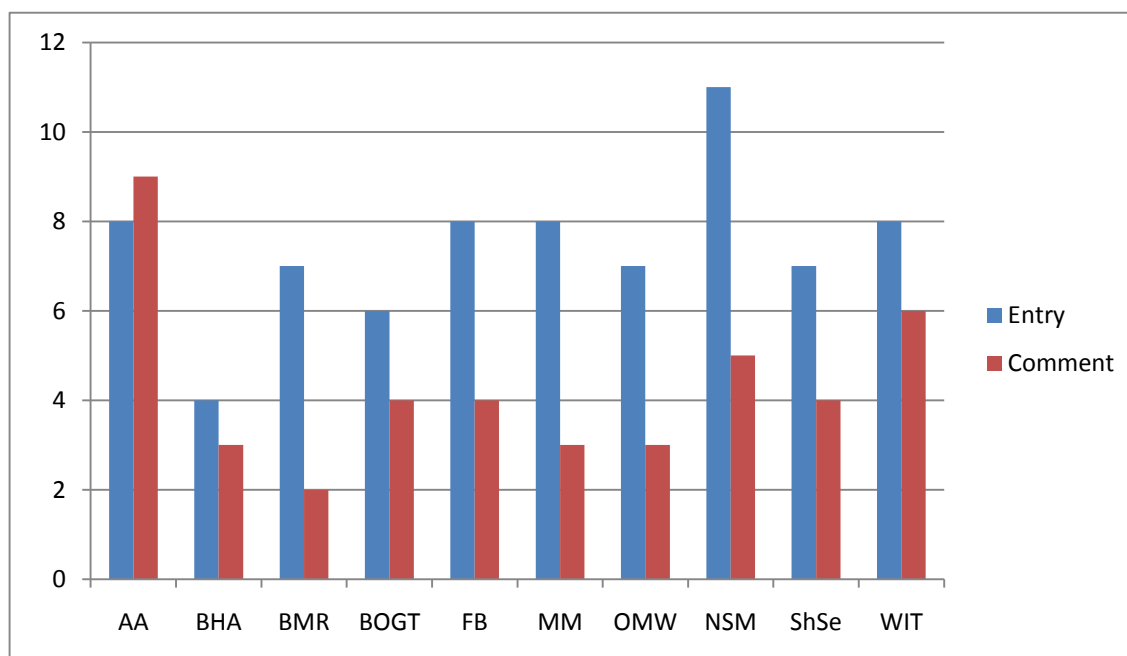


Table 36: Adversative Conjunction in Weblog Entries and Comments

Table 36 portrays the frequencies for this category across the AWC weblog entries and comments. It clearly indicates a considerable decrease of adversative

conjunctions from the entry to the comment sections. The latter is the only weblog featuring higher frequencies in its comments than in its entries. Its high frequency of adversative conjunction is due to an exceptionally frequent usage of one particular adversative conjunction: *but* (55 h/w). This finding suggests that weblog entries exhibit more pronounced narrative dimension in weblog entries than in weblog comments. However, in most weblogs, entry and comment frequencies are not far apart which indicates that weblog comments are equally similar to narrative text types.

The weblog *All Adither* (AA) is the only weblog which shows a higher frequency for adversative conjunctions in comments than in entries. The majority of conjunctive hits can be retraced to the user comments related to two particular entries. In one, the blogger reflects about her “poignant regret” about not having given money to a homeless person (AA/090708/swumcsmd). This entry evokes a plethora of weblog comments, most of which elaborate on possible options of responding to the situation at hand. These comments explore the assets and drawbacks of handouts to homeless people, and they concurrently make frequent use of adversative conjunctions to conjoin their often opposing thoughts and evaluations.

The following list of concordances comprises a short sample of contrasting or concessive contributions from these comments:

7	ness, choice or not. And that always freaked me. <b>But</b> stopping, in your car, to give out spare
8	thanks. I rarely give homeless people money, <b>but</b> when I don't...I think about it for a while
9	the minority here, and maybe I'm a naive bumpkin, <b>but</b> I always give out cash if I have some.
10	Rough topic. Hard to know what to do there, <b>but</b> I tend to err on the side of taking the
11	I don't think handouts really solve the problem, <b>but</b> in the absence of real solutions, at least
12	eal. I prefer food to money as something to give, <b>but</b> that's just me. I remember reading an
13	pidemic of homeless people who have jobs, cars... <b>but</b> no place to live because of the current
14	rogram! Getting involved can be HEART WRENCHING, <b>but</b> also very, very rewarding. It can
15	I give as much as I can. I'm not that religious, <b>but</b> that is definitely a time when I think, "What
16	e's another phrase that comes to mind ... "There, <b>but</b> for the grace of God, go I". I was
17	I admit I was more generous before I was laid off <b>but</b> I give what I can, when I can. I prefer to
19	e to help someone who really needs it when I can, <b>but</b> I rarely carry cash AT ALL anymore.
20	en I can, <b>but</b> I rarely carry cash AT ALL anymore. <b>But</b> I have a pretty zero-tolerance policy for
21	O's and #6 on your list of DONT's can go together <b>but</b> be on different lists!
22	TW, Special K Bars? Not such great hangover food. <b>But</b> alas, it was all I could rummage from

(AA/090708/swumcsmd)

In another comment, the blogger talks about her son's food allergies. A lot of users responded to this entry, expressing their understanding for the son's situation and his struggle to cope with the impairment. They continue to give advice on what the

blogger could do in order to facilitate the life of her child. The first and the second part of the comment are often connected by an adversative conjunction. Consider the following exemplary concordances:

27	, I can't imagine. I would be a wreck), but also to know that your dream of fairly normal eat
28	just a bite, etc. Ack. It gets very old.) But I want you to know that school's take food allerg
29	-free zone. It will annoying and scary, but with the help of the school nurse, I bet you will
30	the epi-pen. I hate being *those people*, but stupid ass food allergies have made it imposs
31	I replied to you on the FEAST board but I just wanted to give you a cheesy (oops, dairy-f
32	you because of the variety of allergies, but will your son have a reaction just being near dai
33	ergy in one. I'm hoping that it'll abate, but at the same time, we eat weird in my house beca
34	. I can't remember the book she read, but this one o
35	I was an intensely sick child for quite a while, but was never denied such a thing as food.
36	ive... I used to cheat on the dairy stuff myself, but no longer worth it AT ALL... Besides just
37	I outgrow this, and that things will get better - but in the meantime, it's worth it to keep Fruit
38	schools are much more aware than they used to be, but the parents still need to be working
39	ow that they can have peanut butter for breakfast but they must still be in their PJs. We scrub
40	und... I don't have anything productive to add but I just wanted to say that I'm sorry. I have a
41	can't even imagine how frustrating that must be. But I bet his teachers are pretty well versed
42	. sorry i have been away a while - hectic times.. but i am back and plan on doing better and
43	not comparing in any way to Fruit Bat's allergy), but I remember the doctor kept trying him on

(AA/280708/tpwigapoofaa)

The high ratio of adversative conjunctions in the comment section can be seen as the result of a specific topical orientation of the two weblog entries. Both of them trigger ample responses which produce contrasting, evaluative judgements on the part of the users. These responses, in summation, lead to the sudden increase of adversative conjunctions in the comment section. I conclude that weblog entries generally show higher frequencies for adversative conjunctions in weblog entries than in comments. Weblog entries thus show a more pronounced narrative dimension than comments. The weblog AA featured unusually high frequencies of adversative conjunctions in its comment section. I explained this anomaly by recourse to the nature of two specific weblog entries. The latter seemed to induce through their topical orientation various comments which all consisted of frequent adversative conjunctions.

Let us now move on to the discussion of causal and temporal conjunction weblogs. Again, Smith & Frawley (1983:353) report the following results for the average distribution of causal and temporal conjunctions across text genres. Causal conjunctions are more frequent than temporal conjunctions in all text genres but journalism. The low frequency of temporal conjunctions in fiction seems odd for a text genre whose narrative structure depends on the chronologic juncture of events. The

authors explain this apparent “absence of temporal conjunctions” by assuming that “temporal succession” is not expressed by “interclausal connection but by lexical progression [...] or by adverbial marking [...] (Smith & Frawley 1983:366).

	Fiction	Journalism	Religion	Science
Total temporal	15	30	23	7
Total causal	18	16	30	26

Table 37: Temporal and Causal Conjunction across Text Genres (Smith & Frawley 1983:363)

In journalism, temporal markers reappear in greater numbers because journalists “place emphasis on the current, durative aspects of [...] events.” (Smith & Frawley 1983:367). While in fiction, the sequential arrangement of past tense verbs represents temporal sequence, journalistic articles often use the historical present in their account of events. The use of temporal conjunctions therefore serves the purpose of framing the exact time of newsworthy events, which is arguably more important in journalistic texts than in fiction.

Let us now take a look at tables 38 and 39 which illustrate respectively the distribution of causal and temporal conjunction across weblog entry and comment sections. We can see that weblog comments generally exhibit higher frequencies for causal conjunction than entries. The weblog *Ohmyword* (OMW) is the only weblog which exhibits higher frequencies for causal conjunctions in entries than in comments.

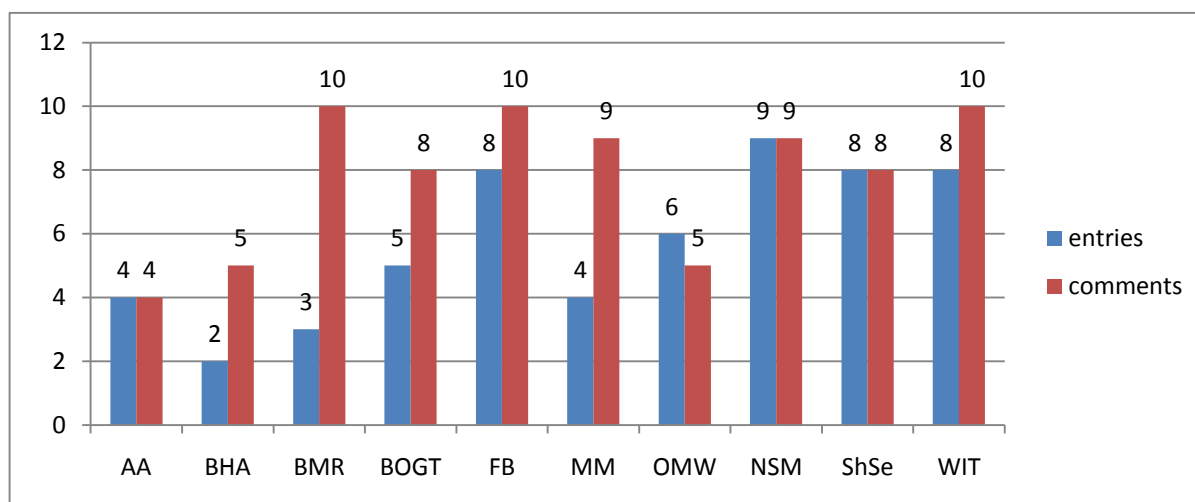


Table 38: Causal Conjunction in Weblog Entries and Comments

Table 39 shows that most weblog comments have higher ratios for temporal conjunction than weblog entries. The difference between the mean score for temporal conjunction in entries and the mean score in comments is substantial. The difference between entries and comments with respect to temporal conjunction is particularly obvious in the two weblogs *Shelli's Sentiments* (ShSe) and *Whatever I think* (WIT).

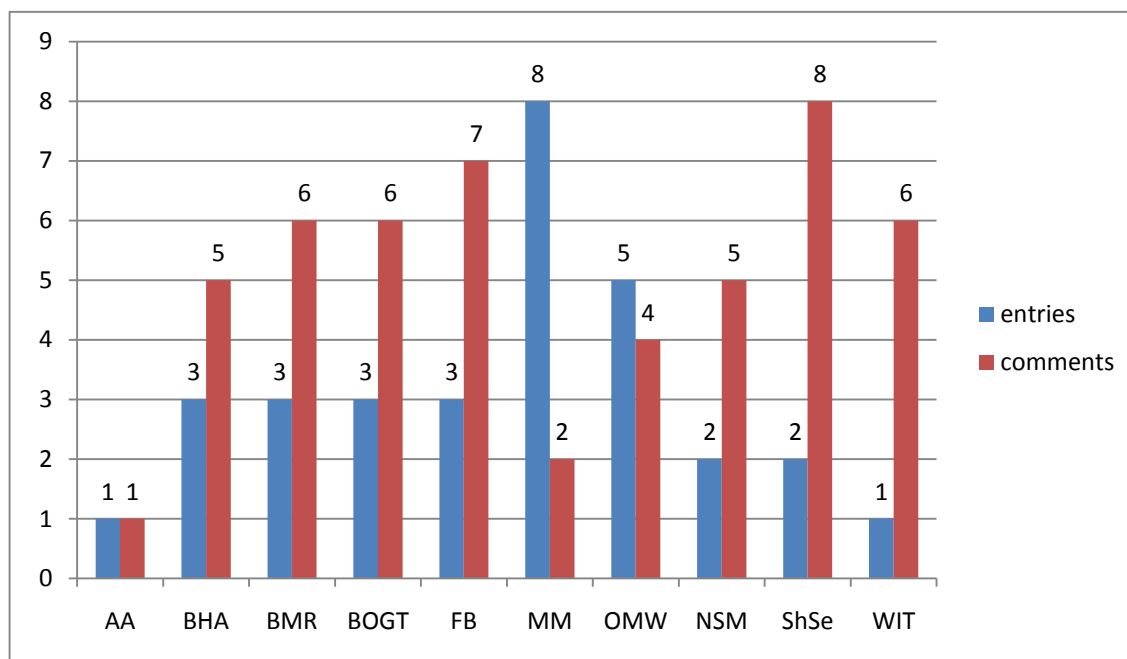


Table 39: Temporal Conjunction in Weblog Entries and Comments

Let us now compare the variation of causal and temporal conjunction for each entry and comment section individually. The following two tables facilitate this task:

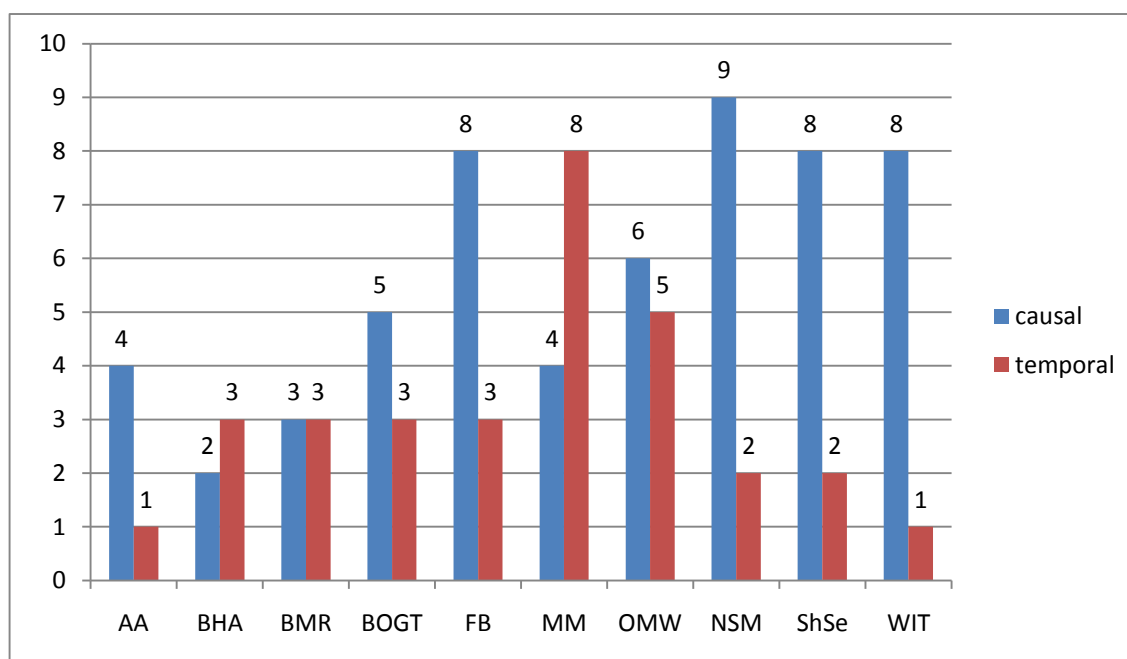


Table 40: Causal and Temporal Conjunction in Weblog Entries



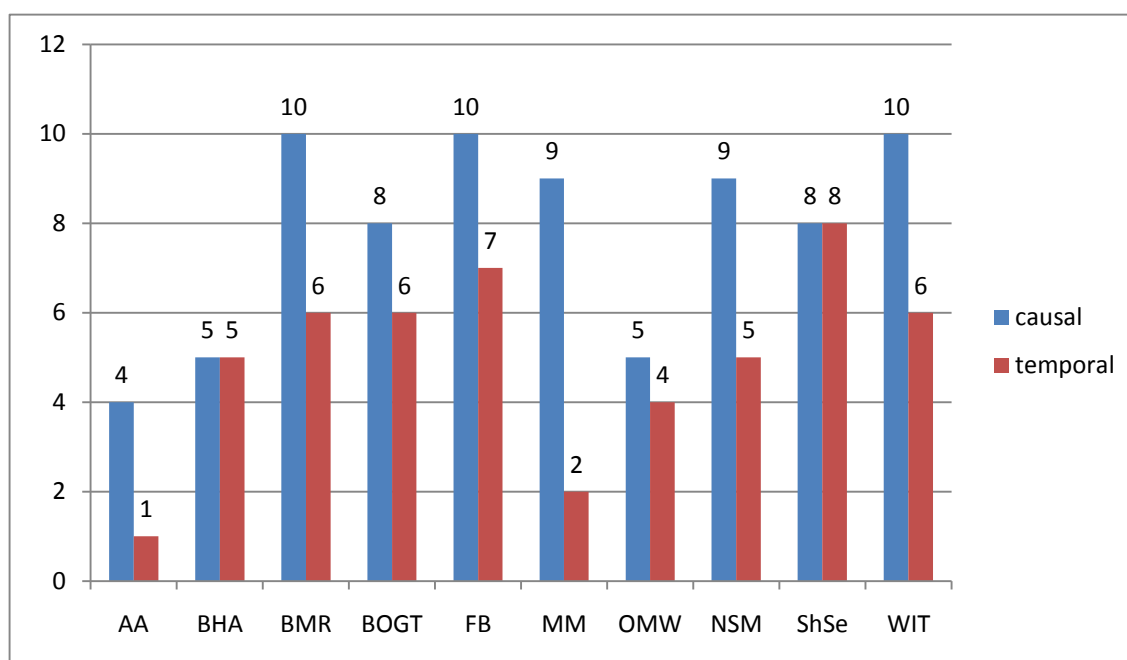


Table 41: Causal and Temporal Conjunction in Weblog Comments

Preliminary to entering the following discussions, it must be stressed that the frequency and distribution of different types of conjunctions can only yield tentative indications of the type of discourse which characterizes a specific weblog section. For instance, we have seen written fiction only includes few numbers of causal and temporal conjunctions. Still, I argue that “the semantics of [conjunctions] give us an excellent insight into the argument and narrative structure of each type of text” (Smith & Frawley 1983:371). Needless to say multiple other studies have repeatedly affirmed this position.<sup>75</sup> It is thus reasonable to assess the level of narrativity for entry and comment section by recourse to their frequent or infrequent use of different types of conjunctions. On the basis of Smith & Frawley’s study on the cross-generic use of conjunctions (1983), I propose the following classification: if two weblogs exhibit high frequencies for additive and adversative conjunctions, they share a distinctive narrative profile. Weblog sections with high temporal frequencies and lower frequencies for causal conjunction show similar cohesive profiles as journalistic articles. In contrast, weblogs with high frequency causal links and low frequency temporal links are likely to be more argumentative rather than narrative in kind. These weblog units seem to be closest to the religious text genre in Smith & Frawley’s study as they included most causal conjunctions of all four text genres

<sup>75</sup> For an exhaustive overview of the previous studies on connectives, consult Halverson (2004:565)

investigated. As Smith & Frawley (1983:369) suggest that there is a “distinct connection between religious and fictive discourse” (Smith & Frawley 1983:369), the latter profile can equally be regarded another cohesive pattern of narrative discourse.

Drawing on this methodological “tool kit”, I shall now ascertain the nature of discourse which governs weblog entry and comment sections. We can capture some important differences between the conjunctive profile of weblog entries and comments in the AWC: both causal and temporal conjunctions are used more frequently in comments than in entries. Especially, the average frequency of temporal conjunction is higher in comments than in entries (3 h/w < 5 h/w). Comments may thus be closer to journalistic discourse than entries, which are more narrative in kind. The difference is accentuated by the importance of temporal connectivity in both journalistic discourse and weblog comments. Both types of discourse make frequent use of the historic present. While journalistic texts seem to value and emphasise the temporal specificity of reportable events, weblog comments highlight the temporal placement of their own weblog contribution in relation to previous entries or comments.

The frequent usage of causal conjunctions in weblog comments points to their complementary narrative dimension which moves them closer to religious texts. The high frequencies for causal and temporal conjunction weblog comments seem somewhat contradictory. Indeed, the text genres investigated by Smith & Frawley (1983) showed either high frequencies for causal *or* temporal conjunctions. Not one text genre exhibited similarly high frequencies for both types of conjunctions. How may we account for this finding? I claim that weblog comments consist of two different sub-types of discourse:

- one is more narrative in kind and thus shows a cohesive profile which is closer to religious text genres (more causal and fewer temporal conjunctions),
- the other one is more argumentative in kind and thus appears to be more similar to journalistic discourse (more temporal conjunctions).

In summation, both types of weblog comments produce high frequencies for both causal and temporal conjunctions. This proves that weblog comments include heterogeneous contributions which alternatively embrace either narrative or argumentative text types.

On the whole, weblog entries show lower frequencies for causal and temporal conjunction than weblog comments. The three weblogs, *NonSoccerMom* (NSM), *Shelli's Sentiments* (ShSe) and *Whatever I think* (WIT) show very high ratios for causal conjunction and much lower ratios for temporal conjunction in their entry sections. I interpreted this finding as an indication of the narrative nature of their discourse. Interestingly, the conjunctive profiles of these weblogs change dramatically in their comment sections which exhibit fairly high frequencies for temporal conjunction, indicative of a more event and time-oriented style of writing. If we compare weblog entries and comments with respect to their general use of conjunctions, we can contend that weblog entries have a more pronounced narrative dimension. Weblog comments, on the other hand, are characterized by a more hybrid type of discourse, which is only partly narrative but may also be argumentative in kind. The only weblog which maintains constant frequencies between entries and comments is the collaborative weblog *All Adither* (AA). It keeps a constant narrative profile between entries and comments which can be traced back to the co-identity of authors and commentators.

Another important finding concerns the weblog, *Mushy's Moochings* (MM). It includes contrasting profiles of its entry and comment section. One finds more temporal than causal links in its entry section while there are more causal than temporal links in its comment section. Its frequency of temporal conjunctions recedes dramatically from eight h/w to two h/w between the entries and the comments. With some justification, we can thus argue that the entry section of this weblog primarily contains argumentative stretches of discourse while the comment section primarily consists of narrative discourse. While there is a general tendency of prevailing narrative discourse in entries, this particular weblog is different. Although the blogger *Mushy* mainly uses his weblog to talk about his past experience, he uses a lot of temporal conjunctions to jump back and forth between various past events he reveals to the reader. He therefore constructs small narrative accounts of his life which are interlaced with one another and connected chronologically by use of temporal conjunctions. This specific style of writing explains the high frequency of temporal conjunctions in the entry section of this weblog.

A third and final interesting point is the general distribution of causal and temporal relations across the weblog comments. There are narrative comments (low frequency of temporal conjunctions) and argumentative comments (high frequency of

temporal relations). Narrative comments occur when users add individual personal stories in the comment sections. Argumentative comments align with users' short personal evaluations of weblog entries. In weblog entries, almost half of the weblogs apply only very few temporal but many adversative conjunctions which is indicative of an elaborate narrative undercurrent in these entries. Only few weblog entries possess more argumentative profiles. Generally speaking, the nature of weblog entries is predominantly narrative, while comment sections include a more variable, dynamic interplay of both narrative and argumentative responses.

### 6.5 Ellipsis in Weblog Entries and Comments

Due to the formal complexity of analysing the cohesive category of *ellipsis* in actual language data, this type of cohesion was excluded from a general quantitative analysis of this study. A consistent analysis of this category would have exceeded the scope of this study. Nonetheless, a rough qualitative analysis of the category was conducted, and instances were found both within as well as between weblog entries and comments. Here are some random examples from different weblog entries in the AWC:

- (60) She was going to make supper but [**she**] needed to needed to get going by 6 p.m.  
(BHA/150708/wd)
  
- (61) I usually lose money ([I] never [**lose**] more [**money**] than \$20)  
(FB/020708/bce)
  
- (62) If thats not a good excuse I don't know what is [**a good excuse**]  
(FB/030708/mbw)
  
- (63) The last one is when Scott and I borrowed bikes and [**Scott and I**] toured the resort, there were so many hills, so we ended up walking our bikes a lot.  
(FB/070708/crhp3)
  
- (64) [...] and although the building is completely vacant, [**the building**] makes more money than any other building in New York City!

(MM/010708/tsa)

(65) [I] Hope she is as proud of me as I am of her.

(MM/040708/hba)

Five weblogs (*All Adither*, *Back Home Again*, *Mushy's Moochings*, *Five Blondes*, *Blog of a Good Time*) were searched for their average occurrence of ellipsis in entry and comment sections. The majority of ellipses in the weblog entries appear to be nominal ellipsis. A quick search of the entry corpus revealed a considerable number of *nominal ellipses* while *verbal* and *clausal ellipses* appear to be much less represented in the entry sections. In fact, there was not a single weblog entry, which contained more than three instances of ellipsis. We can thus argue that ellipsis is only used very moderately in weblog entries. On a general plane, weblog entries therefore exhibit low frequencies for ellipsis. This result can be explained by recourse to the productive setting of weblog entries. The latter are usually written and composed in a well-reflected manner and with great care to detail. This involves a huge amount of planning time. Bloggers are commonly not pressured by time constraints when they create their entries. They are entirely in control of their individual production and publication time. Writing a weblog entry and uploading it to the Internet are two different processes. This is obviously different in conversational discourse settings where interlocutors automatically produce contributions and simultaneously engage in a communicative exchange. Ellipses are linguistic features which accelerate discourse production. On this basis, it has been stressed repeatedly in previous research (cf. Hoey 2005, Martin 1992) that ellipsis is "especially the property of conversation" (Tanskanen 2006:53). It is thus not surprising that, given the specific productive settings of weblog entries, we would find low frequencies for ellipsis in the entry sections. This result reaffirms the previous empirical findings on *reference*, *substitution* and *conjunction* in the entry sections; all of which can be aligned with the monologic rather than spoken cohesive profile.

In contrast to the entry sections, ellipsis is much more frequent in weblog comments. In the five selective weblogs, a rough search elicited numerous instances of nominal, verbal and clausal ellipsis. Here is a random list of some examples found across the different comment sections:

(66) Yours **[audio file]** I would put on my iPod. The earth's **[audio file]**? Nah **[I would not put on my iPod]**.

(AA/020708/ens)

(67) Hey, how do I get me one of those **[pieces of seaglass]**?

(AA/050708/stp)

(68) Maybe we can get together some time and **[we]** make a dress.

(AA/050708/stp)

(69) One thing that appeals to me about sea glass is how happy I can be made by it – **[it is]** much more affordable than other treasures available for \$19.99 all around us.

(AA/050708/stp)

(70) We're headed for the Delaware shore next weekend and **[we're headed for the Delaware shore]** then again for a week in August so I guess I shouldn't complain, but I will **[complain]** anyway.

(AA/050708/stp)

(71) I pretty much tend to give out money. Sometimes **[I give out]** food.

(AA/090708/swumcsmd)

Note that ellipsis in comment sections sometimes refers to discourse elements which are retrievable from the corresponding weblog entry (see examples (66) and (67)). More commonly, however, ellipsis addresses referents in the comment itself. Again, nominal ellipsis was used more prominently than clausal or verbal ellipsis. Nevertheless, considerable numbers of verbal and clausal ellipsis could be identified in the AWC comment sub-corpus. The omission of the first person or second person pronouns (*I* and *you*) at the beginning of comments are a characteristic trait of weblog comments in the data. The following examples illustrate this fairly common feature of comment writing across the different weblogs:

(72) **[You have]** Got the straw spreader removed, eh? It does look wet

(BHA/020708/wh2008)

- (73) Jeffro – [I] Just baled up 5 more loads today. Very heavy and wet.  
(BHA/020708/wh2008)
- (74) Jerry  
[I am] Glad you got home safe. It was sure a lot of fun and I'm really glad I finally got to meet you and Nora.  
(BHA/070708/bfn)
- (75) [You have a] nice profile there - I must second what gunn said. I'm sure this was intentional...  
(BOGT/030708/jn)
- (76) [You are] Doing great so far Lauren!  
(FB/130708/ayip)
- (77) [I] Am LOVING the pics Lau! [I am loving] Especially that one of Kate swinging, [that one of Kate swinging is] definately my fave.  
(FB/130708/ayip)

In contrast to the scarce numbers of ellipsis found in weblog entries, there are numerous instances of ellipsis in weblog comments. It is interesting to note that a high frequency of ellipsis per comment seems to coincide with its low average word length. Hence, comments with fewer sentences are more likely to exhibit instances of ellipsis than longer comments. We can suppose that shorter comments are probably written much more quickly and involve less planning time. I surmise on this basis that the planning time of contributions thus directly imposes on the respective frequency of ellipsis. Less planning time potentially leads to an economical use of lexical means which, in turn, furthers the frequency of ellipsis. We must, however, bear in mind that there are also quite long comments in each of the weblogs investigated. While approximately one half of the weblog comments has an average word length of one to three sentences, the other half exceeds this limit. We can thus expect the frequency of ellipsis in weblog comments to be lower than in two-party and three-party conversations. More importantly, the frequency of ellipsis in weblog comments

is considerably higher than in weblog entries. I argue, however, that this does not imply that weblog comments are conversational, hence dialogic in nature. It only suggests that the planning time involved in creating some of the comments was shorter than in others. I argue that it is not only the interactive dimension of speaker-change and turn-taking which evokes high frequencies for ellipsis. Moreover, it is quite plausible that the time constraints of online language production and reception in conversational settings increase the ratio of ellipses in discourse. Interestingly, there is no actual media-related time limit to the production or reception of weblog comments. Similar to weblog entries, users have full control over their individual production and publication time. Nevertheless, a substantial number of users write short comments in a quick and straightforward, possibly less reflected, fashion.

## 6.6 Some Preliminary Results

To conclude this chapter, I will review and present some of the most important findings discovered so far:

- On a quantitative plane, the cohesive profiles of the ten different AWC weblogs are very similar, showing only slight internal variation.
- Weblog comments exhibit high frequencies for reference relations within individual comments (*internal reference*). Reference relations between entries and comments are generally rare. In addition, these relations are mainly limited to the first sentence of a comment. There are very few reference relations between comments. If cross-comment relations are found, they are usually induced and moderated by weblog authors rather than users.
- On the whole, weblog entries and comments have similar cohesive profiles. Both are closer to the cohesive profile of written monologues (*academic prose*) than to the profile of spoken dialogues (*face-to-face conversation*).
- The general frequencies for comparative reference and substitution are insubstantial in the weblog entry and comment sections. Substitution is *not* more frequent in weblog comments than in entries. This finding again points to the monologic nature of comments. On a qualitative plane, comparative reference seems to bridge intra- and interclausal cohesive relations and thus plays an important part in the construal of weblog cohesion.



- Based on the frequency of conjunctive categories we can infer that weblog entries are argumentative and narrative in kind. Weblog comments show a more pronounced tendency toward a narrative discourse.
- A rough count of ellipsis in four of the ten weblogs revealed a more frequent use of ellipsis in comment than in entry sections. Ellipsis is used more frequently in short comments than in long ones (more than three sentences). If we suppose that short comments are produced more quickly than long comments, then we can argue that production time imposes on the usage of ellipsis. While a considerable number of longer comments could be found in the weblogs investigated, the shorter comments induce a more spontaneous, informal style across the comment sections. While the writing style of comments could thus be qualified as informal or colloquial in some weblogs, the rest of the underlying cohesive tips the balance decisively in favour of a written monologic discourse.

## The Analysis

### Part II: Lexical Cohesion

---

In the previous chapter, we could gain some first insights into the quantitative distribution and qualitative variation of grammatical cohesion in and across personal weblogs. These results revealed that the cohesive profile of personal weblogs is in many respects similar to the cohesive profile of prototypical written monologues, i.e. academic articles. I will test if the result of the previous chapter is sustained by an additional analysis of lexical cohesion. For this purpose, I shall examine the qualitative variation and quantitative distribution of lexical cohesion in personal weblogs. The empirical findings elicited in the course of this analysis will be compared to the cohesive profiles of two-party and three-party conversations (spoken dialogue) and academic articles (written monologue) proposed in a recent study by Tanskanen (2006).<sup>76</sup> The comparison of weblog entries, weblog comments, conversations and academic articles makes for a more exhaustive and detailed analysis of the AWC language data. I shall use the comparative data gained in this process to localise precisely the extent to which weblog discourse (in entries and comments) is similar to classic written monologues and spoken dialogues. Similarities and differences will be interpreted in two principal ways: (a) by recourse to the communicative conditions under which weblog discourse arises and (b) through the qualitative analysis of the purpose and topic which guide the creation of individual discursive contributions. In the end, I shall use the results of the following two chapters to pinpoint the actual position of weblog discourse along the communicative divide of (written) monologue and (spoken) dialogue. The analysis starts with an examination of the distribution of lexical means of cohesion in the AWC weblog entries.

---

<sup>76</sup> The methodological compatibility of Tanskanen's approach enables me to map the results of this study directly onto her findings without any serious practical constraints.

## 7.1 Lexical Cohesion in Weblog Entries

Table 42 gives a general outlook on the variation of lexical cohesion in five different AWC weblog entries. The average distributions of the two main cohesive categories of lexical cohesion, *reiteration* and *collocation*, are illustrated:

	BHA	BMR	BOGT	FB	MM	Conv.1	Conv.2	Conv.3
Reiteration	58	88	97	111	99	126	128	98
Collocation	21	20	24	29	24	15	13	10

Table 42: Reiteration and Collocation Pairs (Weblog Entries vs. Conversations)

As indicated previously, five different weblogs were analysed for their general use of reiteration and collocation, i.e. BHA (Back Home Again), BMR (Beetle's Memories and Ravings), BOGT (Blog of a Good Time), FB (Five Blondes) and MM (Mushy's Moochings). The remaining texts and their respective frequencies, which are enlisted in the table above, were adopted from a recent study on cohesive variation across spoken and written text genres by Tanskanen (2006). The texts *Conv. 1* (Conversation 1) and *Conv. 2* (Conversation 2) represent short excerpts from two-party conversations drawn from the *Lancaster IBM Spoken Corpus* and the *British National Corpus* respectively. *Conv. 3* (Conversation 3) represents a three-party conversatio whose frequencies have equally been drawn from Tanskanen (2006). We can see that between the weblog entries, reiteration frequencies vary considerably.

The weblog MM exhibits twice as many reiteration relations as the weblog BHA. There seems to be less fluctuation with respect to collocation. All weblog entries show fairly similar frequencies in this category. Hence four out of five weblogs present frequencies between twenty and twenty-four h/w. The weblog with the least collocation hits is BOGT with only fifteen h/w. We can now compare these frequencies to the frequencies for reiteration and collocation in two-party and three-party conversations. The contrast between the weblog entries and the two-party conversations is significant. The interlocutors in the two-party conversations evidently use more reiteration pairs than the bloggers in their entries. If we compare the entry sections of the weblog BHA to *Conv. 2*, this contrast is particularly obvious: the conversation contains three times as many reiteration pairs as the weblog entry. It is interesting to note that four out of five weblog entries show more collocation pairs

than the conversations. In addition, the two-party conversations exhibit more collocation pairs than the three-party conversations. How do we account for this difference between two-party and three-party conversations?

Tanskanen remarks that “in an on-line processing production and processing situation, the creation of cohesion is demanding” (Tanskanen 2006:169). In three-party conversations, speakers may not be able to plan their contributions well ahead of time. Consequently, they are more likely to use reiteration pairs than collocation pairs. In two-party conversations, interlocutors appear to have more time to plan their contributions. It appears that the use of reiteration pairs is cognitively less demanding than the use of collocation pairs because time constraints seem to have an impact on their frequency of use. In contrast to conversational settings, bloggers have no time limit at all in the production and publication of their entries and comments. It can be argued that this asynchronous setting of weblogs thus allows for higher frequencies of collocation in weblogs than in conversations.

The frequency differential for collocation pairs between the different conversations is five (10-15). The corresponding range for the weblog entries is nine and thus greater than the one between conversations (15-24). It follows that the variation of collocation pairs is higher between the weblog entries than between the conversations. Nonetheless, the variation between weblog entries is not as high as the variation of reiteration pairs in weblog entries. We can see that the weblog BHA contains 58 h/w while the weblog FB exhibits almost twice as many hits with 111 h/w.

The three-party conversation (*Conv.3*) shows the lowest frequency for collocation pairs of the texts. For instance, the entry section of the weblog MM features more than twice as many collocation pairs than the three-party conversation *Conv. 3*. Four out of five weblog entries have fewer reiteration pairs than the three-party conversation although the difference may not be as significant as the one between entries and two-party conversations. It becomes clear that collocation is used more frequently in weblog entries than in two-party and three-party conversations.

In the following, I will compare the frequencies for reiteration and collocation in weblog entries to the frequencies for reiteration and collocation in written monologues:

	BHA	BMR	BOGT	FB	MM	Article1	Article2	Article3	Article4
Reiteration	58	88	97	111	99	77	78	73	83
Collocation	21	20	24	29	24	16	14	16	14

Table 43: Reiteration and Collocation Pairs (Weblog Entries vs. Academic Articles)

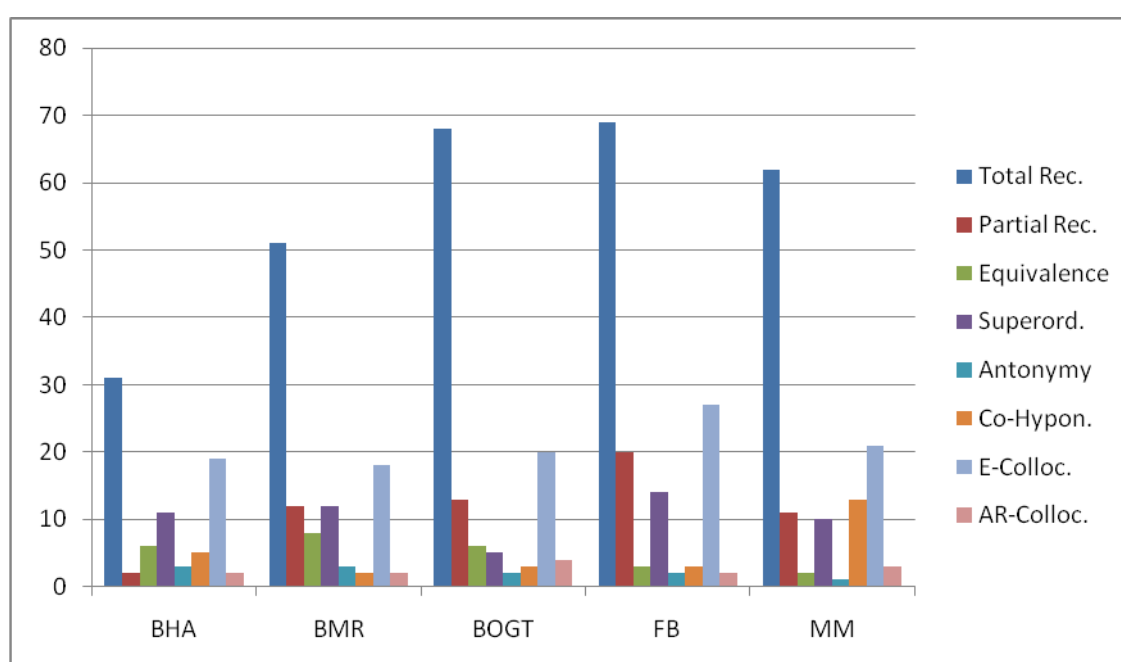
The reiteration and collocation frequencies for texts *article1-article4* are again based on data proposed by Tanskanen (2006). They represent text excerpts drawn from various scientific journals and magazines written and composed for the academic community and can serve as paradigmatic examples for the written monologue. In table 43, it is apparent that the average frequencies for reiteration in the weblog entries and the academic articles are not much different. The weblog entries have a mean score of 91 h/w, whereas the academic articles exhibit a mean score of 78 h/w. The difference amounts to 13 h/w. In contrast, the mean score for conversations are 117 h/w. The difference between the weblog entries and the conversations is thus 26 h/w. Hence there are twice as many reiteration pairs between the conversations and the entries as between the academic articles and the entries.

The frequency for collocation pairs is slightly higher in weblog entries than in academic articles. Across the academic articles, collocation frequencies are stable (14-16 h/w). Again, there is less difference between collocation in weblog entries and academic articles than between weblog entries and conversations. On average, weblog entries are closer to written monologues than to spoken dialogues.

It is fairly plausible to assume that the specific communicative conditions which underlie the composition of weblog entries are responsible for these first observations. The asynchronous communicative setting of weblog entries is, in many respects, similar to the communicative conditions which preset the creation of academic articles. Both types of discourse exhibit the written rather than the spoken mode of communication. They are composed in an asynchronous communicative setting. On-line text production and processing in weblogs is relatively free from temporal or spatial restrictions. This creative liberty of space and time is likely to influence text production. The weblog entries in the AWC generally show a high level of linguistic and thematic complexity which suggests that a considerable amount of planning and reflection was invested in their composition. As most bloggers do not wish to disappoint their target readership, they are careful in selecting the topic they

wish to write and in considering how they wish it to be conveyed (although bloggers may not admit to this freely). As indicated before, the considerable planning time involved in composing entries is equally reflected by the frequent use of collocation pairs.

This first general evaluation of reiteration and collocation relations in weblog entries enables us to examine more closely the various types of lexical relations involved in the realization of both reiteration and collocation pairs. The results obtained from the analysis of lexical categories in AWC weblog entries and comments are shown in table 44.



	BHA	BMR	BOGT	FB	MM
Total Recurrence	31	51	68	69	62
Partial Recurrence	2	12	13	20	11
Equivalence	6	8	6	3	2
Superordinate	11	12	5	14	10
Antonymy	3	3	2	2	1
Co-Hyponymy	5	2	3	3	13
E-Collocation	19	18	20	27	21
AR-Collocation	2	2	4	2	3

Table 44: Lexical Cohesion in Weblog Entries

The most striking result to be inferred from table 44 is the high frequency for *total recurrence* relations in all of the weblogs. Four out of five weblogs contain three times as many instances of total recurrence as the second most frequent reiteration

category. This apparent salience of *total recurrence* pairs across the weblog entries indicates that blogger's make heavy use of interpersonal pronouns *I*, *you* and *we*, which I have classified as instances of *total recurrence* in this study.<sup>77</sup> In the same vein, Puschmann (forthcoming) informs us that a frequent use of interpersonal pronouns is common not only in personal weblogs. For instance, he reports of corporate weblogs that these essentially "follow the conventions of personal blogs closely" (Puschmann, forthcoming). This fact can be seen as another indication for the pervasive influence of the personal weblog genre within the blogosphere. It is possible that personal weblogs might export some of their salient formal characteristics to other weblog genres. With a view to the high use of total recurrence pairs, we have reason to believe that a high frequency of interpersonal pronouns is, in fact, a general, characteristic of weblog writing.

In weblog entries, the first person singular pronoun is conventionally identifiable as the author and publisher of a weblog. As such, it is applied consistently in almost every weblog entry of the AWC. The reference of first person singular pronouns in weblog entries is therefore easy to resolve and stable. In contrast, the identification of second person singular and plural pronouns is more demanding on the reader as pronominal reference may shift from one addressee to the next. Consequently, *person deixis* reference in weblogs is at once stable (first personal singular) and variable (second personal singular/plural); it generates a hybrid characteristic of weblogs which Puschmann (forthcoming) coins *blog deixis*.<sup>78</sup>

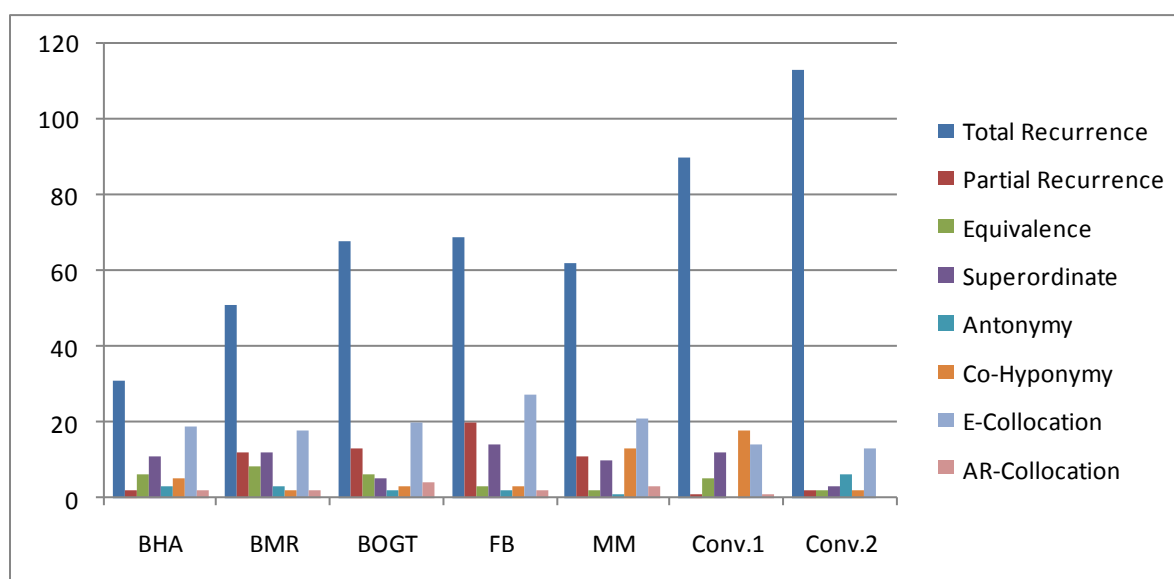
Let us now return to the additional types of lexical cohesion in table 45. The second largest category of lexical cohesion is *elaborative collocation*, followed by *partial recurrence* and *superordinate*. The fifth highest category of lexical cohesion is *equivalence*, followed by *co-hyponymy*, *antonymy* and *activity-related collocation*. The individual profiles of each weblog are quite similar, bar a few minor exceptions. For instance, the frequency for total recurrence in the weblog *Mushy's Moochings* (MM) seems unusually high compared to the other weblogs. It equally features the highest ratio of *co-hyponymy* of all the weblog entries.

---

<sup>77</sup> Tanskanen (2006) also subsumes interpersonal pronouns under the category of *total recurrence*. The normalized frequencies she provides in her investigation thus remain fully compatible with the quantitative findings of the present study.

<sup>78</sup> Following Bühler (1999/1934), I define *deixis* as the prototypical linguistic means of pointing which acquire meaning by recourse to a given context of situation. Speakers commonly use deictic expressions to locate their subjective position in regard to a given place and time of utterance.

In order to put the frequencies for these cohesive categories into perspective, I shall compare them to the frequencies for lexical cohesion in face-to-face conversations and academic articles. I will begin by comparing the AWC weblog entries with the standardized frequencies for lexical cohesion in two-party conversations (cf. table 45). Evidently, the frequency for *total recurrence* is much higher in the two party conversations than in the weblog entries. This finding aligns with the results of *reiteration* and *collocation* discussed previously. *Reiteration* was found to be much more prevalent in conversations than in the weblog entries and academic articles. I also pointed out that collocation pairs are more frequent in our weblog entries than in the conversations. This result is affirmed by the average frequencies for elaborative and activity-related collocation in table 45. We can add that most collocation pairs both in the weblog entries and the conversations are classified as *elaborative collocations*. There are only few activity-related collocations in the corpus.



	Total Rec.	Partial Rec.	Equivalence	Superord.	Antonymy	Co-Hypon.	E-Colloc.	AR-Colloc.
BHA	31	2	6	11	3	5	19	2
BMR	51	12	8	12	3	2	18	2
BOGT	68	13	6	5	2	3	20	4
FB	69	20	3	14	2	3	27	2
MM	62	11	2	10	1	13	21	3
Conv.1	90	1	5	12	0	18	14	1
Conv.2	113	2	2	3	6	2	13	0

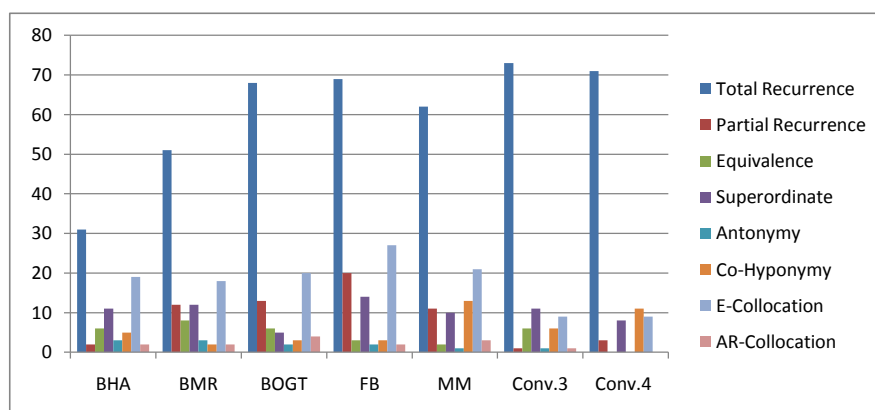
Table 45: Cohesive Variation of Lexical Cohesion in Weblog Entries and Two-Party Conversations

Nonetheless, there are slightly more activity-related collocations in the weblog entries than in the two-party conversations. The dissimilar frequencies for total recurrence



and elaborative collocation are the most salient contrasts between the cohesive profiles of the weblog entries and the conversations. The weblog entries have higher frequencies for *partial recurrence* than the two-party conversations. *Superordinate* relations are used consistently across all the texts in equal numbers, with the exception of the weblog *Blog of a Good Time* (BOGT) and *Conversation 2* (Conv.2). Whereas Conv.2 shows a low frequency for *co-hyponymy*, Conv.1 exhibits the highest frequency for *co-hyponymy* of all the texts. The frequencies for *co-hyponymy* oscillate from one conversation to another. *Equivalence* could be elicited in all texts but appeared in minor numbers only. There are more instances of equivalence in the weblog entries than in the conversations but the quantitative difference is inconclusive.

Now, we can compare the frequencies for different cohesive categories in weblog entries and three-party conversations. Table 46 illustrates the respective frequencies for our AWC weblog entries and the two three-party conversations from Tanskanen (2006). The conversations have higher frequencies for *total recurrence* than the weblog entries but this time the difference is less extreme.



	Total Rec.	Partial Rec.	Equivalence	Superord.	Antonymy	Co-Hypon.	E-Colloc.	AR-Colloc.
BHA	31	2	6	11	3	5	19	2
BMR	51	12	8	12	3	2	18	2
BOGT	68	13	6	5	2	3	20	4
FB	69	20	3	14	2	3	27	2
MM	62	11	2	10	1	13	21	3
Conv.3	73	1	6	11	1	6	9	1
Conv.4	71	3	0	8	0	11	9	0

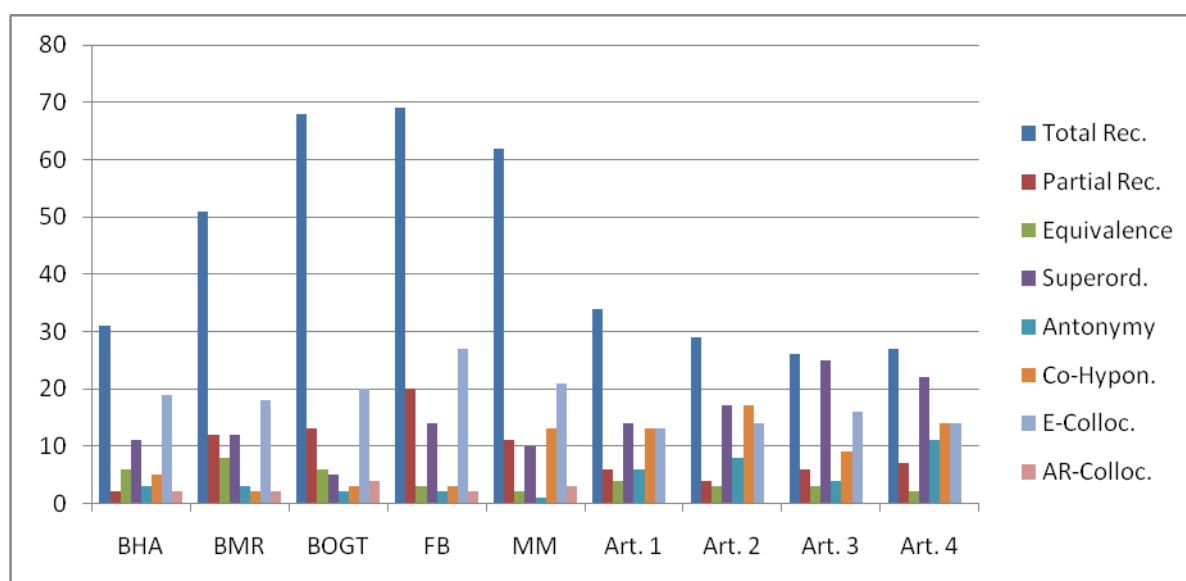
Table 46: Cohesive Variation of Lexical Cohesion in Weblog Entries and Three-Party Conversations

The three-party conversations have lower frequencies for *elaborative collocation* than the two-party conversations. With respect to elaborate collocation, the disparity between weblog entries and conversations has increased. Similar to the two-party conversations, the three-party conversations show only few instances of *partial recurrence*. The weblog entries have generally higher frequencies for this category than both the two-and the three-party conversations. It is interesting to note that the co-hyponymy seems to be applied consistently in the three-party conversations and that the frequency of co-hyponymy is higher in the three-party conversations than in the two-party conversations. In weblog entries, however, co-hyponymy is used very rarely.

If we take the results of the previous cohesive categories in two-party conversations as a blue print, we can make the following inferences for the discourse in weblog entries:

- weblog entries exhibit lower frequencies for *total recurrence* than conversations.
- weblog entries exhibit higher frequencies for *partial recurrence* than conversations.
- weblog entries exhibit higher frequencies for *elaborate* and *activity-related collocation* than conversations.
- weblog entries exhibit lower frequencies for *co-hyponymy* than conversations.

Before proceeding to the interpretation of these findings, I will first compare the normalized frequencies for lexical cohesion in the weblog entries with the corresponding frequencies in the academic articles adopted from Tanskanen (2006). The following table illustrates the distribution of reiteration and collocation pairs in weblog entries and in four different academic articles.



	Total Rec.	Partial Rec.	Equivalence	Superord.	Antonymy	Co-Hypon.	E-Colloc.	AR-Colloc.
BHA	31	2	6	11	3	5	19	2
BMR	51	12	8	12	3	2	18	2
BOGT	68	13	6	5	2	3	20	4
FB	69	20	3	14	2	3	27	2
MM	62	11	2	10	1	13	21	3
Art. 1	34	6	4	14	6	13	13	0
Art. 2	29	4	3	17	8	17	14	0
Art. 3	26	6	3	25	4	9	16	0
Art. 4	27	7	2	22	11	14	14	0

Table 47: Cohesive Variation of Lexical Cohesion in Weblog Entries and Academic Articles

We can see that the frequency for *total recurrence* is higher in the weblog entries than in the academic articles. Still, the difference is not as great as the one between weblog entries and two-party conversations. As we can see in table 48, the mean score for total recurrence in academic articles is 29 h/w, the mean score for conversations is 87 h/w and the mean score for weblog entries is 56 h/w. There is a difference of 27 h/w between weblog entries and academic articles, and a difference of 31 h/w between the average frequencies for weblog entries and conversations. The weblog entries find their position between the spoken dialogue and the written monologue. If all interpersonal pronouns are factored out of the analysis, the mean score for total recurrence drops to 31 h/w. In other words, the frequency for total recurrence is almost reduced by half.

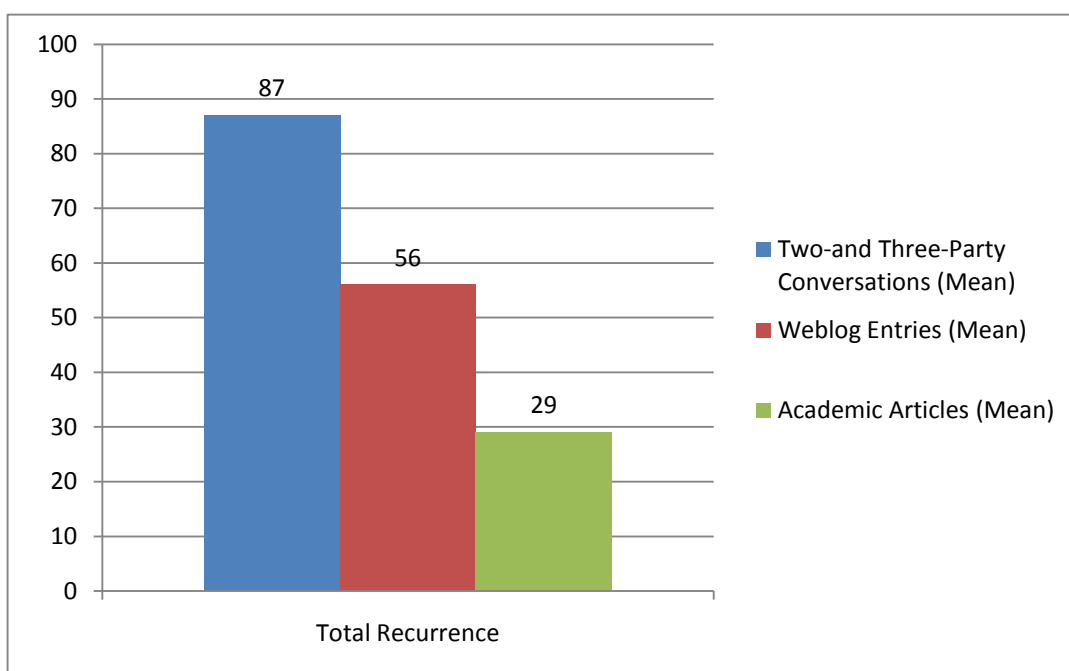


Table 48: Mean Scores for Total Recurrence (incl. Interpersonal Pronouns)

This confirms the immense impact of interpersonal pronouns on the average ratio of total recurrence in the weblogs. Once the interpersonal pronouns are deleted from the cohesive profile, the mean score for total recurrence in the weblog entries is very similar to the one of the written monologues (cf. table 49). In conclusion, the weblog entries show more instances of total recurrence than the academic articles; yet, and importantly, they show fewer hits than two-party conversations. If I bracket off all interpersonal pronouns, there is almost no difference between academic articles and weblog entries (29 vs. 31 h/w) at all. In contrast, the difference between entries and spoken dialogue is considerable (31 h/w vs 87 h/w).

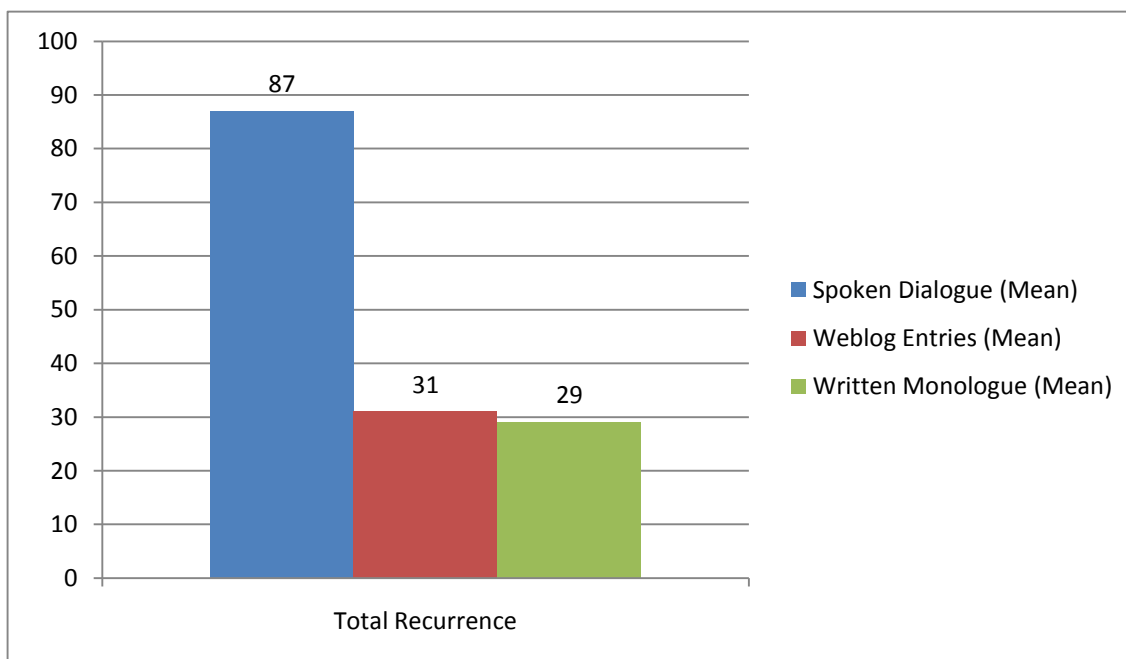


Table 49: Mean Scores for Total Recurrence (Weblog Entries Without Interpersonal Pronouns)

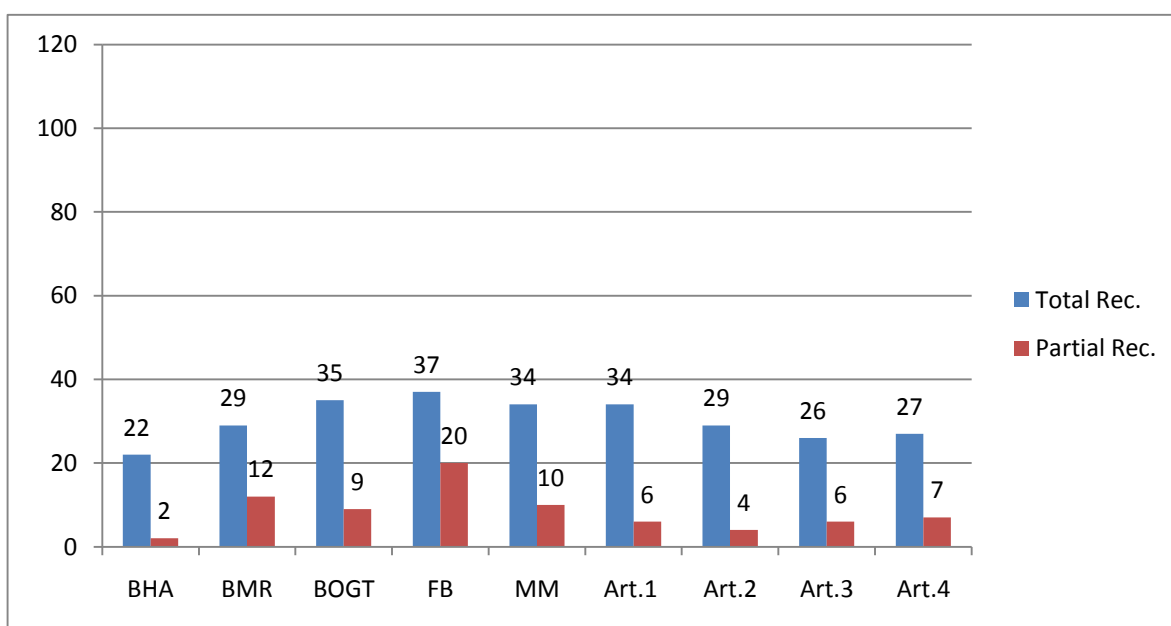


Table 50: Total and Partial Recurrence in Weblog Entries (- interpersonal pronouns)

Table 50 confirms that most interpersonal pronouns which enter into a cohesive relation adhere to the category of total recurrence, such as *you - you*, *I - I*, *we - we*. Whereas the blue bars are considerably smaller than in the previous tables, the red bars indicating partial recurrence have largely maintained their original size. Hence, interpersonal pronouns which enter into a cohesive relation of partial recurrence, e.g. *you - your*, *he - him*, etc., are much less frequent in the entries.

Admittedly, the frequency for total recurrence in the conversations (spoken dialogue) would be expected to decrease if we deleted all interpersonal pronouns. Nonetheless, the reduction is likely to be much less extreme than the one described for the weblog entries. Furthermore, the similarity between the reduced profile of total recurrence in weblog entries and the profile of the academic articles is strikingly similar. The overall quantitative difference of 27 h/w between the use of total recurrence in the weblog entries and the academic articles can therefore be retraced exclusively to the bloggers' frequent use of interpersonal pronouns. If we only compare the mean scores for total recurrence in the two-party conversations and the weblog entries, the difference is even higher than between the weblog entries and both the two- and three-party conversations (cf. table 48 and 51).

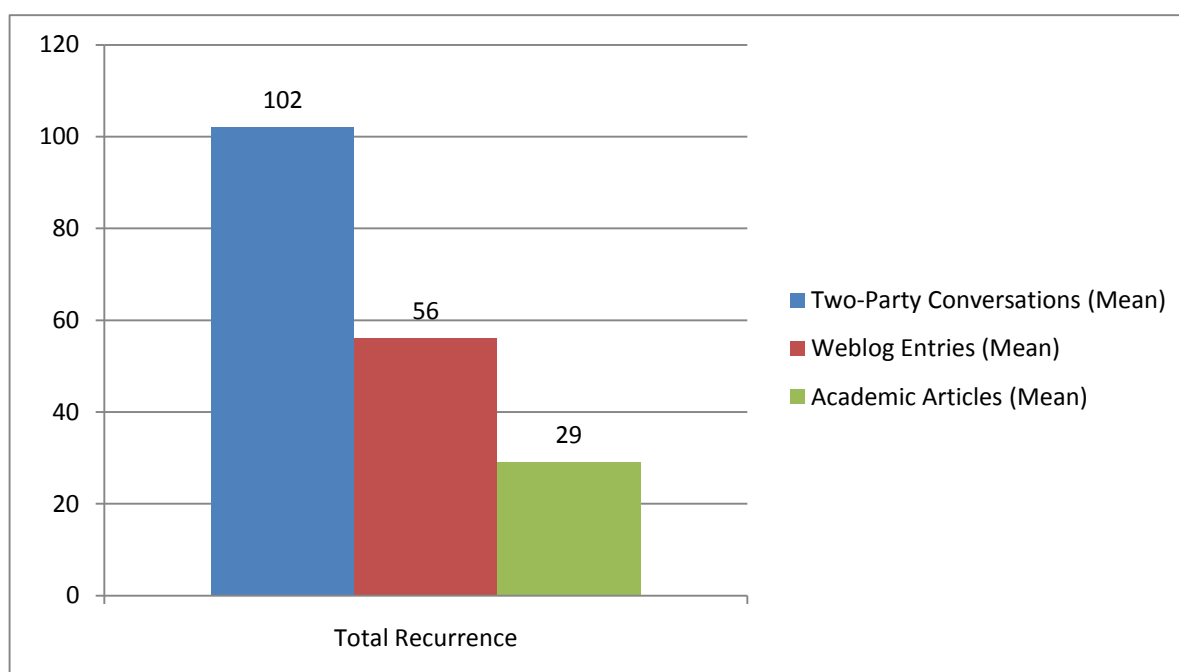


Table 51: Mean Score for Total Recurrence in Two-Party Conversations, Entries and Academic Articles (incl. interpersonal pronouns)

The mean score for the two-party conversations is 103 h/w, for the weblog entries is 56 h/w and for the academic articles 29 h/w. There is a difference of 27 h/w between the weblog entries and the academic articles. In contrast, the difference between the weblog entries and the two-party conversations is twice as high, with a differential of 52 h/w. Hence the two-party conversations feature approximately twice as many hits as the weblog entries (cf. table 51).

Table 52a illustrates this difference with respect to the individual scores for total recurrence in each weblog entry. It is evident that the two two-party conversations have higher scores than the weblog entries (conversations = orange and light blue column). The dashed lines indicate the respective mean scores for total recurrence in conversations (blue), weblog entries (green) and academic articles (red). Evidently, the difference between the weblog entries and the academic articles is less pronounced than the difference between the weblog entries and the two-party conversations.

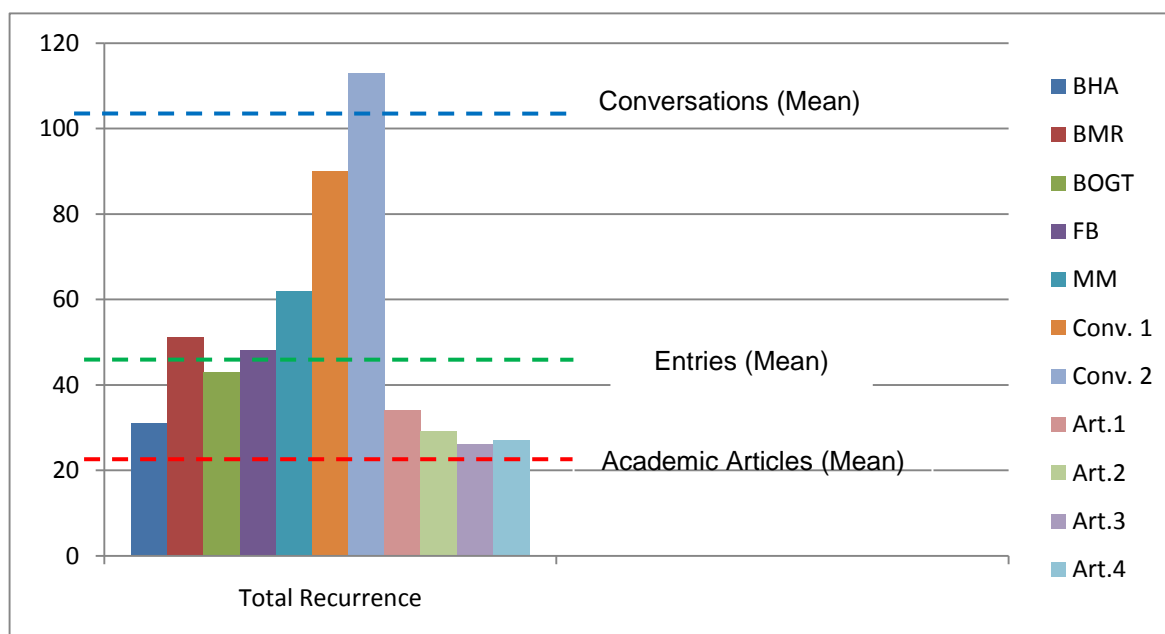


Table 52a: Total Recurrence across Weblog Entries, Conversations and Acad.Articles (Mean Scores)

Another important finding is revealed in table 52b. The mean score for partial recurrence in the weblog entries is much higher than in the conversations or the academic articles. We can detect a gradual increase of partial recurrence from conversations to academic articles and finally to weblog entries. It is interesting to note that, once again, the mean score for partial recurrence in the academic articles is thus closer to the one in weblog entries than the mean score of conversations. I take this as another indication of the fact that weblog entries bear a much closer resemblance to written monologues than to spoken dialogues.

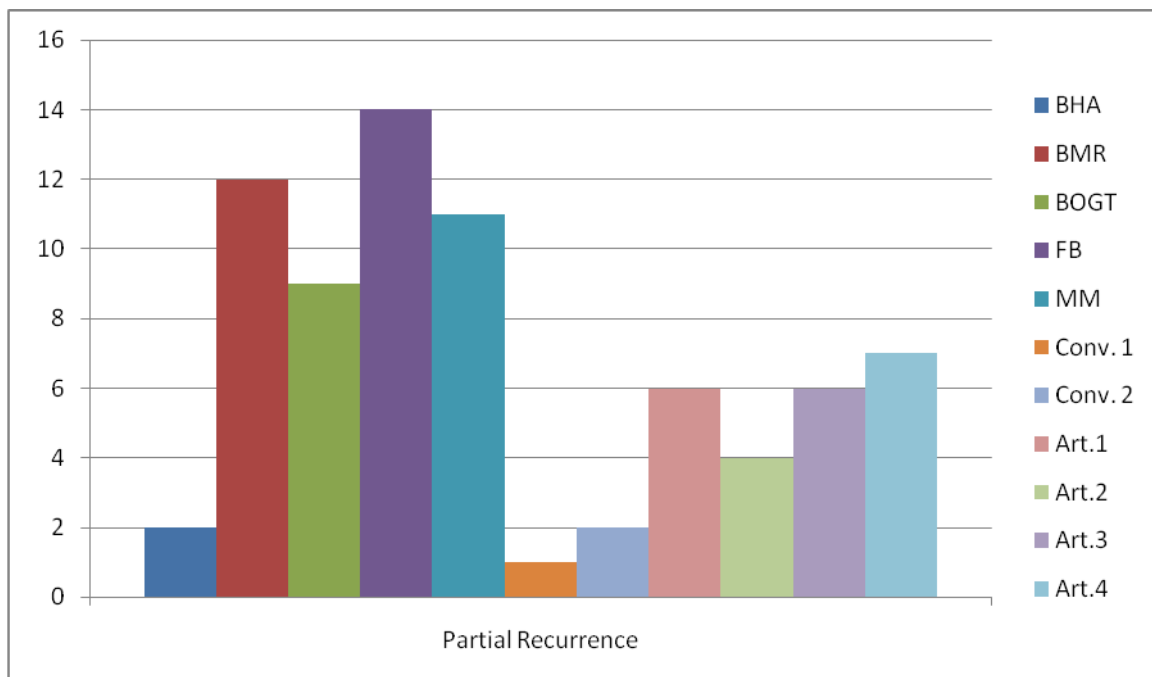


Table 52b: Partial Recurrence across Weblog Entries, Conversations and Acad.Articles

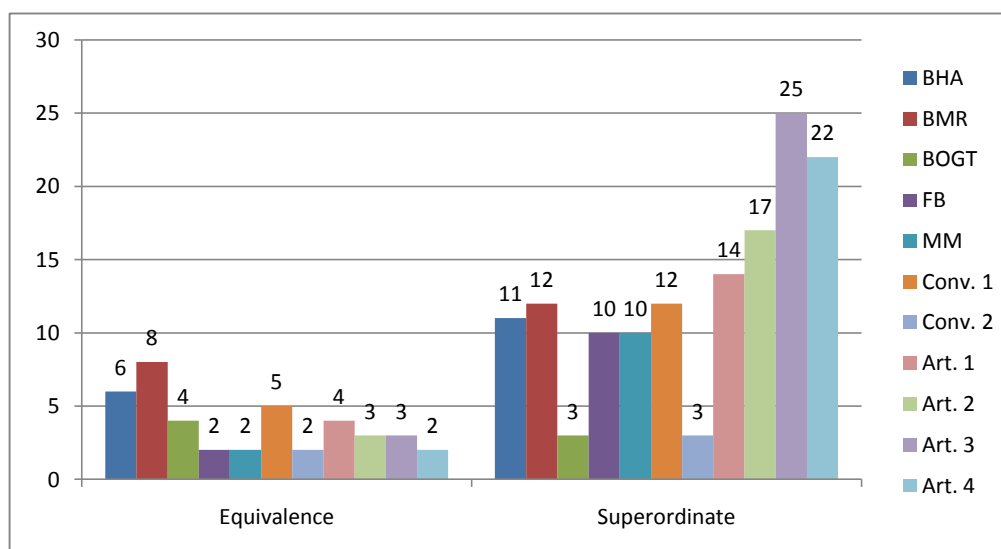


Table 53: Equivalence and Superordinate in Weblog Entries, Conversations and Academic Articles

The frequencies for *equivalence* are fairly consistent across weblog entries, conversations and academic articles. There is no substantial difference between these three groups of texts. The academic articles have higher ratios for *superordinate* relations than the conversations and the weblog entries. The mean score of *superordinate* relations in the academic articles is 17 h/w, the mean score in the weblog entries is 9 h/w and the mean score in the conversations is 8 h/w. In this category, the weblog entries are actually closer to the two-party conversations than to the academic articles.



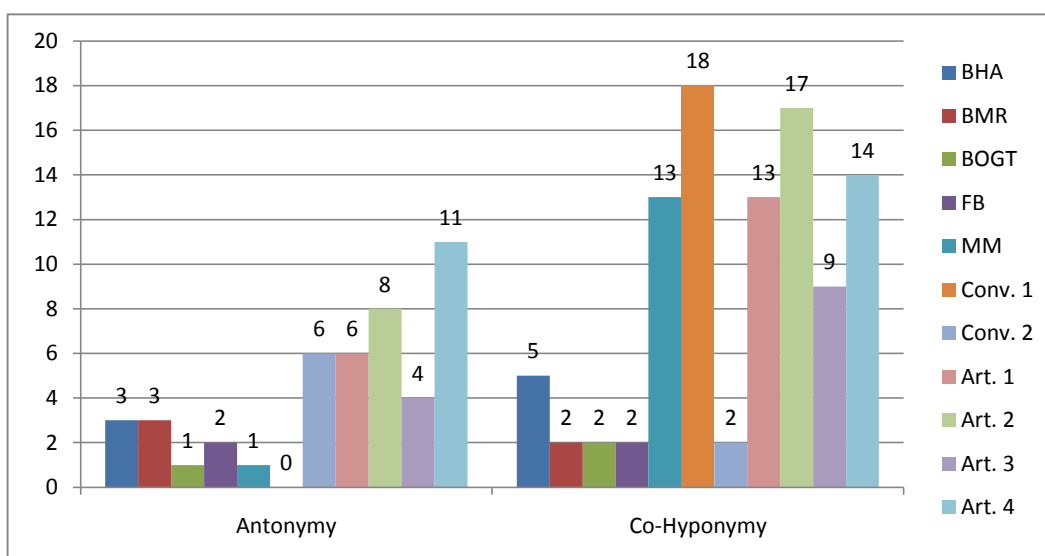


Table 54: Antonymy, Co-Hyponymy in Weblog Entries, Two-Party Conversations and Acad. Articles

The categories *antonymy* and *co-hyponymy* are used quite inconsistently in the two two-party conversations. This makes it difficult to establish a precise cohesive profile for the conversational text genre in these two categories. Nonetheless, we can see that the average number of antonyms and co-hyponyms in the conversation and the academic articles are higher than in the weblog entries. The high frequency of antonymy and co-hyponymy in the academic articles is a result of the argumentative style central to scientific discourse. In academic articles, examples or empirical findings are usually compared and contrasted with each other. This “comparative” style accounts for the high mean scores for both antonymy and co-hyponymy in this text genre. The more personal writing style in the weblogs differs in theme and purpose from the academic articles thus manifesting fewer instances for antonymy and co-hyponymy.

It would be interesting to compare fictional texts to personal weblogs as one would expect both written text genres to have similar average ratios for both cohesive categories. Hence, it is likely that frequencies for antonymy and co-hyponymy are heavily contingent upon the topical orientation of text genres as well as their conventionalised style.<sup>79</sup> Generally speaking, the frequency for antonyms and co-hyponyms in the weblog entries and conversations vary considerably. The topic-

<sup>79</sup> I suspect that the stark contrast between the frequencies for both cohesive categories in *Conv. 1* and *Conv. 2* is related their two very different conversational topics. It seems possible that one of these topics would lend itself more readily to the use of antonyms and co-hyponyms than the other.

dependency of antonymy and co-hyponymy does not actually allow for any concise discrimination of generic differences between conversations and weblog entries, apart from the general tendencies mentioned above.

I shall now conclude the analysis of weblog entries with a short review of the empirical findings discussed so far:

1. The most salient category of lexical cohesion in the weblog entries is *total recurrence*. Almost half of all the hits analysed in this category were instances of interpersonal pronouns (*I*, *you*, and *we*). If we include interpersonal pronouns in our analysis of mean scores for weblog entries, the latter assume a bridging position between the two- and three party conversations and the written academic articles. However, with respect to the mean scores of reiteration pairs, weblog entries are closer to written monologues than to spoken dialogues. If we compare only the two-party conversations (prototypical spoken dialogue) to the weblog entries and the academic articles (prototypical written monologue), the weblog entries clearly have much closer similarities with the written monologue than with the spoken dialogue. Furthermore, I factored out all interpersonal pronouns from the list of total recurrence relations in the weblog entries. As a result, the mean score for total recurrence in the weblog entries drops considerably and is almost on par with the one in the academic articles. The difference between the frequencies for total recurrence in the weblog entries and the academic articles can therefore be solely retraced to the excess of first personal singular and plural pronouns in the weblogs.
2. The weblog entries display partial recurrence more frequently than the conversations or academic articles. In contrast, the latter two genres have higher frequencies for co-hyponymy than the weblog entries. These differences can partly be explained through variety of purpose and related writing conventions. Scientific argumentative texts seem to privilege co-hyponymy and antonymy relations at the expense of other cohesive relations and in contrast to the entries and conversations.
3. The weblog entries exhibit a higher frequency for collocation than the conversations and the academic articles. On a quantitative plane, the mean score of collocation across the weblog entries is much closer to the mean

score of collocation for the academic articles than to the mean score for the conversations. The similitude of the cohesive profiles of the weblog entries and the academic articles can be largely explained by recourse to the communicative conditions, which they seem to share. The written mode of production enables an asynchronous interactive setting which frees interlocutors from on-line production and processing constraints. It is likely to account for the higher frequencies of collocation pairs. An excess in planning time for the creation of weblog entries and academic articles seems to provide an explanation for the higher numbers in them, hence we may suspect that the application of collocation is more cognitively demanding than the use of reiteration relations.

4. On average, the weblog entries show lower frequencies for superordinate, antonymy and co-hyponymy than the conversation and the academic articles. The academic articles have the highest mean score for all of these categories. However, the difference between the mean scores of the weblog entries and mean scores (of all these categories) for the weblog entries and the mean scores for the academic articles is not as high as the one elicited for total recurrence between the weblog entries and the two-party conversations. In fact, at no time does the differential between the weblog entries and the academic articles exceed 10 h/w. Consequently, the difference between weblog entries and academic articles is rather subtle and can be considered as insignificant.

## 7.2 Lexical Cohesion in Weblog Comments

In the following, I shall proceed to analyse the distribution of lexical cohesion in our weblog comments. Table 55 presents the average distribution of reiteration and collocation pairs for the weblog comments and the conversations in our corpus:

	<b>BHA</b>	<b>BMR</b>	<b>BOGT</b>	<b>FB</b>	<b>MM</b>	<b>Conv.1</b>	<b>Conv.2</b>	<b>Conv.3</b>
Reiteration	119	117	119	81	84	126	128	98
Collocation	23	33	21	15	23	15	13	10

Table 55: Reiteration and Collocation Pairs (Weblog Comments vs. Conversations)

The most striking observation we can make from table 55 is the high frequency for reiteration pairs in the weblog comments. Whereas the mean score for reiteration pairs in the weblog entries is 91h/w, the mean score in the weblog comments is 104 h/w. The reiteration pairs have thus increased by 13 h/w from the weblog entries to the comments. This finding seems to position the weblog comments closer to the conversations which presented a mean score of 117 h/w (for both two-party and three-party conversations). However, if we consider only the mean score for reiteration pairs in the two-party conversations and leave out the three-party conversations, the mean score for conversations increases to 127 h/w. It is feasible to concentrate primarily on two-party conversations here because this text genre is usually assumed to be the classic prototype for spoken dialogues. If we compare the weblog comments and the two-party conversations (rather than the two-and three-party conversations), their counts for total recurrence differ by 23 h/w. As table 56 illustrates, the average ratio for reiteration pairs in academic articles is 78 h/w. There is a difference of 26 h/w between the mean score of reiteration in the weblog comments and the corresponding mean score in the academic articles. The differential for reiteration pairs between the weblog entries and the two-party conversations on the one hand and the one between the weblog entries and the academic articles on the other are quite similar (23 h/w vs. 26 h/w). Hence, we can infer from these preliminary findings that weblog comments again assume a bridging position between spoken dialogue and written monologue with respect to its use of reiteration pairs. The difference between weblog comments and conversations is significant.

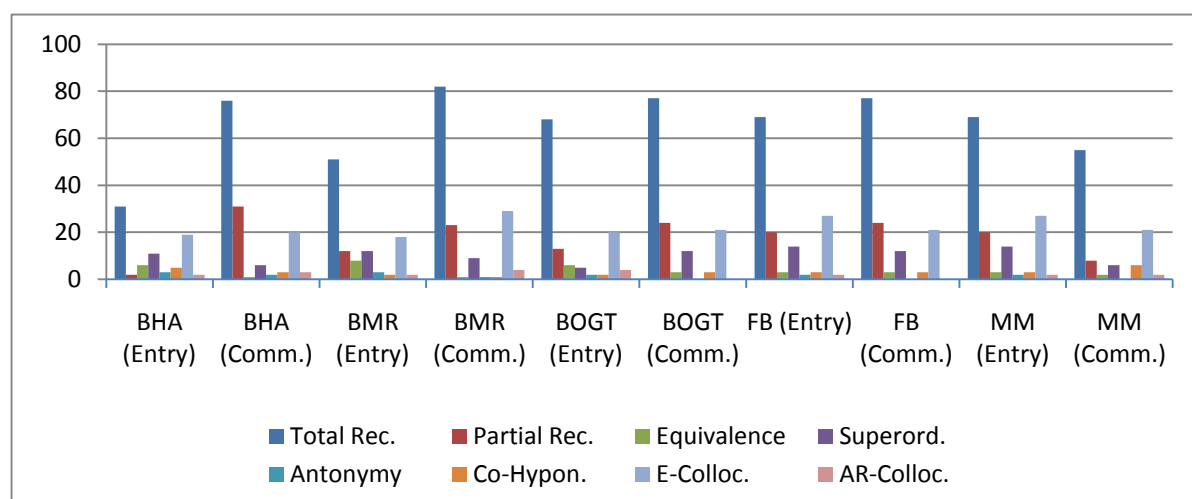
The mean score for collocation pairs in the weblog comments is almost identical with the corresponding score in the weblog entries (24 h/w vs. 23 h/w). Both the weblog entries and the comments therefore feature more collocation pairs than the conversations and the academic articles. More importantly, the difference between the mean score for collocation in the weblog comments and the mean score for collocation in the academic articles is less substantial than the corresponding difference between the weblog comments and the two-party conversation. Again, I suggest that in regard to collocation, the weblog comments bear a closer resemblance to the written monologue than to the spoken dialogue.

	BHA	BMR	BOGT	FB	MM	Article1	Article2	Article3	Article4
--	-----	-----	------	----	----	----------	----------	----------	----------

Reiteration	119	117	119	81	84	77	78	73	83
Collocation	23	33	21	15	23	16	14	16	14

Table 56: Reiteration and Collocation Pairs (Weblog Comments vs. Academic Articles)

Let us now turn to assessing the most substantial changes between the cohesive profile of lexical cohesion for weblog entries and weblog comments.



	Total Rec.	Partial Rec.	Equivalence	Superord.	Antonymy	Co-Hypon.	E-Colloc.	AR-Colloc.
BHA (Entry)	31	2	6	11	3	5	19	2
BHA (Comm.)	76	31	1	6	2	3	20	3
BMR (Entry)	51	12	8	12	3	2	18	2
BMR (Comm.)	82	23	1	9	1	1	29	4
BOGT (Entry)	68	13	6	5	2	2	20	4
BOGT (Comm.)	77	24	3	12	0	3	21	0
FB (Entry)	69	20	3	14	2	3	27	2
FB (Comm.)	77	24	3	12	0	3	21	0
MM (Entry)	69	20	3	14	2	3	27	2
MM (Comm.)	55	8	2	6	0	6	21	2

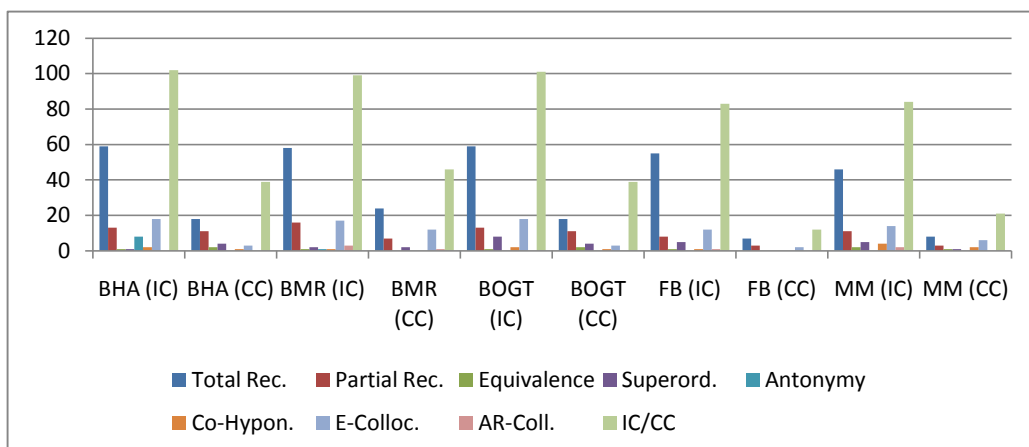
Tables 57 and 58: Lexical Cohesion in Weblog Entries and Comments

Table 57 captures the most elementary changes between the frequencies of lexical cohesion in our weblog entries and comments. It is apparent that total recurrence is more frequent in the weblog comments than in the entries, with the notable exception of the weblog *Mushy's Moochings* (MM). In addition, partial recurrence is generally more frequent in the comments than in the entries. The frequencies for equivalence are slightly higher for the weblog entries than for the weblog comments although the difference is not significant. All weblogs exhibit similar ratios for *superordinate* in both the entries and the comments, with minor fluctuations. There are only few instances of antonymy and co-hyponymy in both entry and comment sections. Some weblogs increase, others decrease their individual score for collocation from the entries to the comments. On the whole, however, these frequency shifts are not substantial. In our

weblog entries and comments, almost all the hits for collocation are generated by elaborative collocation, showing only few instances of activity-related collocation.

It is obvious from the above findings that the main difference between weblog entries and comments is their use of total and partial recurrence: comment section generally include more instances of total and partial recurrence than their corresponding entry sections. I will investigate more carefully the nature of lexical relations in weblog comments in order to account for the significant increase in recurrence relations from entries to comments.

In weblog entries, cohesive pairs are always created by a single author only. In contrast, the analysis of weblog comments revealed two different types of lexical relations, i.e. relations within single comments (IC= inside the comment) and relations between comments (CC= cross-comment). The frequencies for lexical cohesion presented above included both IC and CC cohesive relations. In other words, cohesive relations were analysed within and across individual comments. Let us look into the distribution of IC and CC relations within the different comment sections. Table 58 illustrates that the profiles for IC-relations in weblog comments are remarkably similar across the weblogs. All of the weblog comments feature roughly the same amount of pairs for total recurrence (46-60 h/w), partial recurrence (11-13 h/w) and elaborative collocation (12-18 h/w). The sum total of IC pairs is equally stable across the different weblog comments with 80-100 h/w. In addition, we can observe that the frequencies for CC-relations in the comments are much lower than the corresponding IC-relations. The CC-relations of total recurrence only amount to one-fourth of all lexical relations in each weblog. In contrast, the frequency of partial recurrence is stable between IC and CC-relations. Equivalence and superordinate relations are low both within and across the comments but superordinate relations generally show slightly higher ratios for IC-relations. Antonymy and co-hyponymy relations are equally low within and across the comments. Finally, there are more IC-relations than CC-relations of elaborative and activity-related collocation in the weblog comments.



	Total Rec.	Partial Rec.	Equivalence	Superord.	Antonymy	Co-Hypon.	E-Colloc.	AR-Coll.	IC/CC
BHA (IC)	59	13	1	1	8	2	18	0	102
BHA (CC)	18	11	2	4	0	1	3	0	39
BMR (IC)	58	16	1	2	1	1	17	3	99
BMR (CC)	24	7	0	2	0	0	12	1	46
BOGT (IC)	59	13	1	8	0	2	18	0	101
BOGT (CC)	18	11	2	4	0	1	3	0	39
FB (IC)	55	8	1	5	0	1	12	1	83
FB (CC)	7	3	0	0	0	0	2	0	12
MM (IC)	46	11	2	5	0	4	14	2	84
MM (CC)	8	3	1	1	0	2	6	0	21

Tables 58 and 59: IC vs. CC-Relations in Weblog Comments

The results therefore indicate that IC-relations are stable across all the weblogs, and that more than half of all lexical relations could be identified as IC-relations. The most significant difference between IC-relations and CC-relations in the comment sections concerns the category of total recurrence. Other cohesive categories are not affected as much. The majority of hits in the class of total recurrence are IC-relations. As total recurrence equally represents to most frequent cohesive category in our weblog comments; this difference is significant for our interpretation. Although few instances of total and partial recurrence between comments were elicited, there were no substantial numbers for any of the other remaining cohesive categories. The low frequencies for CC-relations (with regard to lexical cohesion) thus confirm the low frequencies for cross-comment reference with regard to grammatical cohesion as observed in the previous chapter. The cohesive interaction between the weblog comments therefore seems to be largely restricted both on a grammatical and lexical plane. I infer that most weblog comments are grammatically and lexically self-contained units of discourse. They mainly establish and progress cohesive relations within their textual boundaries, i.e. intertextual references are only very moderately

applied between comments and entries and between comments themselves the occur to an even lesser degree.

In contrast, lexical relations between comments and entries do exist. However, the referents of lexical relations in the comments are often quite difficult to discern in their related entry units. As a matter of fact, in the analysis potential referents were found to be dispersed all across a particular weblog entry. Long weblog entries thus had to be read entirely in order to identify the possible referents for lexical items in a comment section. Example (78) illustrates the difficulty of referent resolution across text units. The comment includes the lexical items: *the 80's*, *the oldies station* and *radio*, which all refer to related lexical items in a preceding weblog entry. These co-referents are *Eighties*, *radio* and *station*, and they are located at the beginning rather than the end of the weblog entry. The respective clause which is addressed by the commentator is reproduced in example (78). In order to resolve the reference of the lexical item, readers first must read the entire weblog entry to establish the underlying lexical bonds between the comment and its entry:

(78) Comment: Yeah, **the 80's**...when I listen to **the "oldies" station** on the **radio** I crank it up just to watch my daughters cringe! I consider it payback for listening to their music...

(BHA/180708/tsdmwh)

Entry: [...] Any time there is an Eighties Weekend on a radio station, this one will definitely be played. [...]

(BHA/180708/tsdmwh)

As referents can be located in the beginning, middle or end of weblog entries, the resolution of possible lexical (and grammatical) cohesive relations can become quite cumbersome for users, especially if the word length of weblog entries is high. Long weblog entries commonly include more than two identity chains, i.e. they address multiple topics. This fact further complicates the search for potential intertextual referents. As a result, users may fail to resolve the intertextual cohesive relation. In addition, some of the comments in the AWC did not refer to the entry to which they were assigned but rather responded to earlier entries posted by the same blogger. We may thus conclude that while most weblog comments entertain some sort of lexical relations to their related entry section, the individual elements and types of relations are often difficult to identify for users.



This broad cohesive scope of entry-comment relations makes the elicitation and quantification of cohesive hits extremely problematic. In the previous chapter, the cohesive scope of referential relations was extended to the entire weblog entry. This approach seemed only appropriate for grammatical cohesion. We should remember that personal and demonstrative anaphora could only be interpreted by recourse to their co-referent which had to be identified either in the comment itself (internal relation), a preceding comment (cross-comment) or the preceding entry (entry-related). In lexical cohesion, however, the anaphora (a noun, verb or adjective phrase) is interpretable on its own. It does not need the prior identification of its co-referent to be meaningful. Therefore, an extension of the cohesive scope of analysis to the entire weblog entry seems inadequate in the case of lexical cohesion. Apart from this, the broad scope of analysis is likely to produce misleading findings, rendering them incompatible with the other quantitative results of this study.

Let us now return to the interpretation of our comparative results for lexical cohesion in weblog entries and comments. The two categories of total recurrence and partial recurrence showed higher frequencies in the weblog comments than in the entries. The analysis of IC-relations and CC-relations proved that while partial recurrence was produced partly within and partly between comments, total recurrence primarily registered within single comments. How do we explain this general increase of total and partial recurrence in the weblog comments compared to the mean scores for total and partial recurrence in the entry sections? It seems likely that interpersonal pronouns are more frequently used in weblog comments than in weblog entries. The responsive purpose of weblog comments seems to induce a more frequent application of first person and second person singular pronouns (*I* and *you*). To this end, table 60 illustrates the frequencies for total and partial recurrence without interpersonal pronouns in our weblog comments:

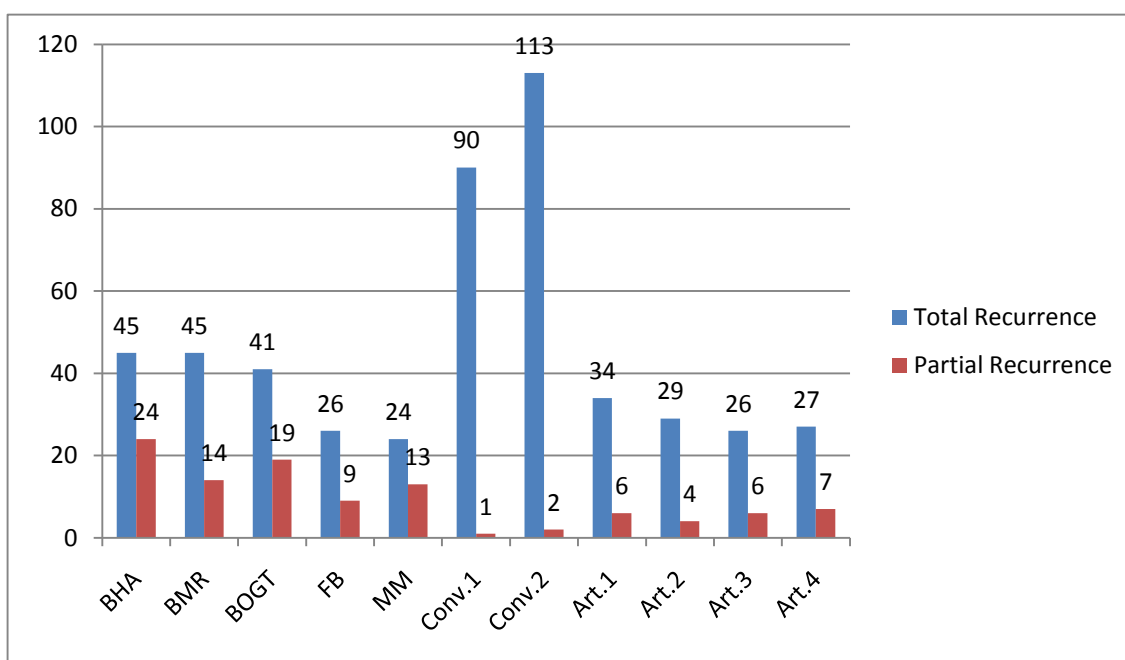


Table 60: Total and Partial Recurrence in Weblog Comments (- interpersonal pronouns)

Table 60 shows that the frequencies for total recurrence are greatly reduced once interpersonal pronouns have been deleted from the analysis. The weblog comments have a mean score of 73 h/w for total recurrence relations with interpersonal pronouns. Without interpersonal pronouns, the mean score for total recurrence in the weblog comments is 36 h/w. As a result of the deletion of interpersonal pronouns from the cohesive profile of comments, the mean score for total recurrence is reduced by half. In other words, half of all the hits elicited for total recurrence in the comment sections could be identified as interpersonal pronouns. As the weblog entries, the cohesive profile for total recurrence in the weblog comments is quite similar to the profile of the academic articles. The quantitative difference between the weblog comments and the academic articles is minor with a differential of only 7 h/w (36 h/w vs. 29 h/w). With a view to the distribution of total recurrence, we can argue that the cohesive texture of the weblog comments is quite similar to the one of the weblog entries. Both include a considerable amount of interpersonal pronouns which are responsible for the high frequencies for total recurrence in both weblog entries and comments. Entries and comments feature higher frequencies for partial recurrence than the two-party conversations and the academic articles.

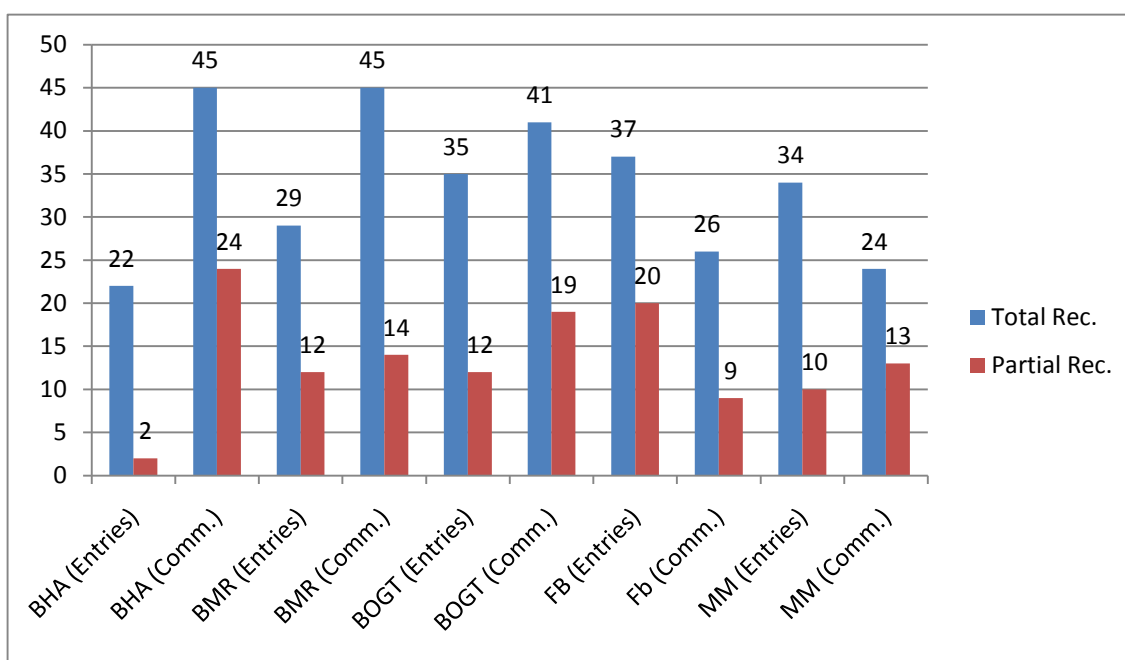


Table 61: Total and Partial Recurrence in Weblog Entries and Comments (- interpersonal pronouns)

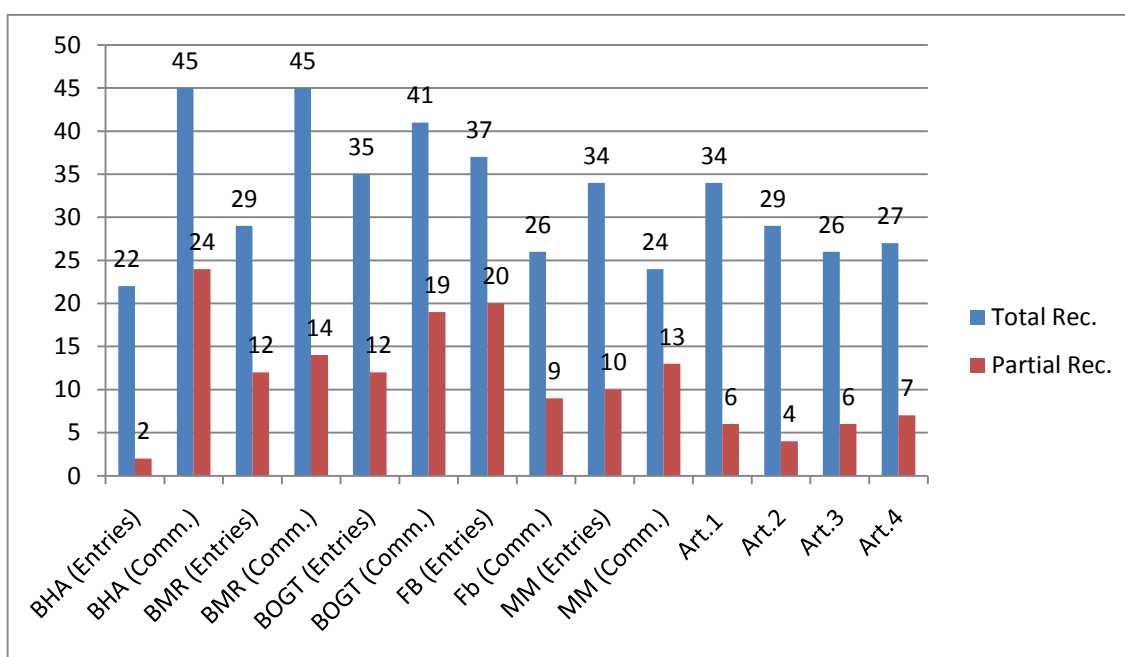


Table 62: Total and Partial Recurrence in Entries, Comments and Acad. Articles (- interpersonal pr.)

Once all interpersonal pronouns are withdrawn from the weblog entries and comments, the frequencies for total and partial recurrence are surprisingly consistent (31 h/w vs. 36 h/w) between the weblogs. The profiles are not only strikingly similar; both are equally close to the mean score for total recurrence in the academic articles (29 h/w). Face-to-face conversations also include a considerable amount of inter-

personal pronouns, and it is likely that these are responsible for the high frequencies for total recurrence in the dialogical profiles of Conv.1 and Conv. 2. Despite this fact, the correspondence of the frequencies between the weblog entries and the comments (- interpersonal pronouns) and the academic articles is remarkable. It suggests that the profiles of both types of weblog discourse bear closer similarities to the cohesive profile of academic articles than to the one of two-party conversations. In addition, the high ratio for collocation in both the weblog entries and comments is closer to the frequency of collocation in the academic articles than in the two-party conversations.

Even if I include interpersonal pronouns in our analysis of total recurrence in the weblog entries and comments, the frequencies for both types of text are not as high as the one registered for the two-party conversations which represents the prototype of dialogical communication.

### **7.3 Preliminary Results**

On the basis of the previous comparative analysis, I can conclude the following results from our analysis of lexical cohesion in and across the weblog entries and comments:

- Weblog entries and comments show similar profiles for lexical cohesion, both have high frequencies for total recurrence and comparatively high frequencies for collocation.
- Weblog comments exhibit a higher mean score for reiteration pairs than weblog entries. Similarly, they feature a higher mean score for total and partial recurrence than weblog entries. This difference can be explained by recourse to the responsive nature of weblog comments which entails a recurrent use of the interpersonal pronouns *I* and *you*. While the first person pronoun singular is used more often in weblog entries, the second person singular is applied more regularly in weblog comments.
- The high frequency of total recurrence in weblog entries and comments can be explained by the frequent occurrence of interpersonal pronouns. More than half of all the instances of total recurrence found in the entry and comment sections could be identified as interpersonal pronouns.
- If all interpersonal pronouns are factored out, the cohesive profiles of both entries and comments are similar to the profile of academic articles. Scientific

writers appear to make more recurrent use of superordinate, antonymy and co-hyponymy relations than weblog authors or users. Nonetheless, the quantitative difference between the text genres is quantitatively inconclusive.

- Both the weblog entries and the comments have similar scores for collocation which mainly consists of elaborative collocation. These scores are higher than the frequencies for collocation in the academic articles and the two-party conversations, yet on a quantitative plane, they are closer to the former than to the latter.
- The use of lexical cohesion in weblog entries and comments is very close to the cohesive profile of the written monologue (academic articles), while it differs much more from the standard cohesive profile of the spoken dialogue (two-party conversations). This finding can be explained by the similarity of communicative settings between weblog entries and comments on the one hand and academic articles on the other. Both types of communications arise under the conditions typical of the written mode of production. They take place in an asynchronous discourse setting and impose few on-line production and reception constraints on their interlocutors. In both types of discourse, interlocutors invest considerable time and effort into the production of their discourse contributions. The recurrent speaker change and the more demanding interactive setting of spoken dialogues induce a cohesive texture which is, in many respects, different from the one elicited from weblogs.
- An important stylistic similarity between two-party conversations and weblog entries and comments is their frequent usage of interpersonal pronouns in them. This simulates a quality of dialogicity in weblogs which the underlying cohesive texture does not sustain. I argue that the mock effect of the use of pronouns is that users are under the illusion that they are actually engaged in a regular conversation when, in fact, the ordinary length and complexity of weblog interaction in personal weblogs is extremely limited (at least in personal weblogs).

## The Interaction: Knowledge and Cohesion

---

### 8.1 From Collocation to Cognition

At the end of this study, I shall now return the interrelation of cohesion and coherence in weblogs. The focus of my attention shall be devoted to a more thorough examination of the cohesive category of collocation to illuminate its special role at the intersection of interlocutors, cognition and discourse. It has been emphasised that weblog entries and comments exhibited a considerable number of collocation relations. I have argued the inclusion of the category of collocation into the frameworks of cohesion used in this study based on assumption that they fulfil a viable cohesive function in the qualitative and quantitative analysis of discourse. As we have seen, differences in the frequency and use of collocation across different text genres reveal important insights on their individual productive and receptive constraints of text genres. I thus agree with Tanskanen (2006:34) who claims that the controversial arguments which have been voiced against the category of collocation are “hardly grounds for excluding the entire category from an analysis of lexical cohesion”. Far from being an elusive analytical tool, I argue that interclausal collocations can be used to bring to the fore a number of intriguing aspects of weblog discourse:

- (a) collocations can be classified to reveal the main topical areas bloggers and users deal with in entries and comments.
- (b) collocations reveal recurring lexical elements through which these topics are established, progressed and negotiated.
- (c) the subjective tinge of collocations connects them to specific types of knowledge necessary to identify a discursive tie as collocational.
- (d) collocations mirror which kinds (or groups) of users are targeted by bloggers and which ones are not addressed. This enables us to attach participation roles to specific groups or regular, frequent or new users.

Let us begin by classifying the different collocation pairs which were elicited across the weblog entries into topical groups. In order to become coherent, each topical group can be connected to cognitive structures through which users assign meaning

to discourse, i.e. *cognitive frames*. I have thus subsumed each group of collocates to the cognitive frame which are lexically induced across the different weblog entries.<sup>80</sup> The most frequent (salient) frames in the AWC weblog entries and the respective collocational pairs allocated to them are summarised in the following:

- **[TECHNOLOGY]** → [PHOTOGRAPHY], [COMPUTER], [INTERNET], [VIDEO]
- **[ENTERTAINMENT]** → [MUSIC], [FILM]
- **[TRAVEL]** → [BUILDINGS], [ROAD/TRAFFIC], [VEHICLE]
- **[KINSHIP]**
- **[INGESTION/FOOD]**
- **[COMMERCIAL]**
- **[TIME]**

These frames operate on different cognitive planes. Frames, such as [TECHNOLOGY], [ENTERTAINMENT] or [TRAVEL] are more abstract than the others. They are thus located on a high level of cognition. They are more complex and abstract than low level frames, such as [PHOTOGRAPHY], [MUSIC] or [BUILDINGS]. The latter, in turn, are more concrete and thus structure and organize the respective high level frame.<sup>81</sup> The arrows indicate the lower level frames complete which, in turn, complete the higher level frames. This classification of frames has been derived from the quantification and topical categorization of collocations. We may now use it to explore how the different cognitive frames impose on the construction and (cognitive and discursive) negotiation of interlocutors.

## 1. [TECHNOLOGY]

The majority of bloggers share an interest for communication technology. This is manifest in the considerable number of entries which are devoted to the discussion of communication media e.g. a photo camera, mobile phone, MP3 player across the weblogs. Users who respond to these “tech-posts” identify themselves as individuals who share specialised knowledge about the technology being discussed with the blogger. The [TECHNOLOGY] frame is frequently evoked across weblogs. I suggest

<sup>80</sup> The designations for frames used in the course of this classification have been largely adopted from Charles Fillmore’s online *framenet* databank which is available on the Internet at: <http://framenet.icsi.berkeley.edu/> (25/09/09).

<sup>81</sup> High level frames are printed in bold letters in the above list. Low level frames are designated by standard font capital letters in brackets.

that entries which induce this frame are directed at tech-savvy Internet users. They are expected to be familiar with web-related terms and concepts such as *voki*, *blog feed*, *flickr*, *facebook*, *widgets*, *code*, etc., whose meanings are not explained within the entries. Rather, bloggers who use such lingo in their writing already presuppose a certain degree of technological knowledge from their readers. Consider the following examples excerpted randomly from different weblog entries:

(79) New **blog template**- or is it?

I am a very happy bunny today.

Can you see any difference in my **blog design**?

When I designed my **blogs** I used a **web design programme**, which was great for the designing part, but didn't produce the best **code** in the world. It left me with lots of problems to overcome, and lots of fiddling about every time I added a **widget**! The worst problem was the method used to **upload** my **template**. I had to **copy** and **paste the code**. This, of course meant that I lost all of my **widgets** and had to start from scratch!

(BMNR/010708/nbtolii)

(80) The Name **Meme**

I'm not a great one for **Meme**'s but I came across this one and thought it was fun. It's another 'Google Search' **meme**. [...]

(BMNR/070708/tnm)

(81) I've made a **flickr account**, and once I get a theme I'll put pictures on it, and put the address up on here so you all can look at my super exciting life...[...] Edit-by the way, the cameras name is Sophie Dallas, because there is an SD in the type name.

(FB/050708/pedfay)

(82) [...] You will notice that I now have a pretty signature on my **posts**. I have also removed **Digg** and **SezWho**, and replaced it with a 'Stumble upon'. At least that's something to mark a new year of **blogging**!

(BMNR/230708/imtb)



## 2. [ENTERTAINMENT]

Another frequent frame, which is triggered by collocation relations in weblog entries, is [ENTERTAINMENT]. Some weblogs regularly introduce, evaluate and discuss music albums, bands, pop stars as well as feature films, directors and film stars. Indeed, four out of five weblogs featured at least one or two entries dedicated to these topics. Again, bloggers usually assume that regular visitors of the weblog share a certain amount of cultural knowledge on English-language films and music. For instance, US film stars, soap operas are expected to be familiar to the user as the following excerpt illustrates:

- (83) Why OK! never should have put Jamie Lynn on their cover [...] On the front cover of this particular issue of OK! is 16-year-old Baby Mama **Jamie Lynn Spears** posing with her newborn. "Being a mom is the best feeling in the world!" she tells us. It's all so wonderful and great! And she tries to convince us that Casey (Baby Daddy), with eyes as big as pie plates and a grimace that's supposed to pass for a grin, is not going to bolt as soon as the spotlight is off the trio [...] Not only does OK! make teen motherhood sound like a dream-come-true, it violently distorts motherhood in general. Even for those of us who are mature, married and relatively balanced, being a mom is frackin' HARD. Motherhood is not all about "hanging pictures in the nursery" and buying cute, little outfits. It's about tar-black shitty diapers and no sleep and snapping at your significant other in the middle of the night and, occasionally, wondering what the hell have I gotten myself into? [...]
- (AA/160708/wonshpjlotc)

In this excerpt, the blogger presupposes a series of facts to be known by the readers for them to be able to contextualize and fully comprehend this weblog entry. These presuppositions refer inter alia to the facts (a) that *Jamie Lynn Spears* is the sister of the popular US-American pop idol Britney Spears and (b) that *OK!* is one of the largest US magazines. The target audience of the above weblog entry is thus revealed as US-American with a dual interest in both celebrity gossip and parenting. On the one hand, bloggers shape their virtual identity by their individual choice of topics and the linguistic means through which these are presented to users. On the other hand, target readerships of particular weblogs can be reconstructed from the

type and degree of knowledge presupposed in the different entries of a weblog. In this vein, let us take a look at another example from a different weblog entry:

(84) FACT: **Tori Spelling** is boring and annoying.

I'm pretty sure **Tori Spelling** has been boring and annoying since being ejected from **Candy Spelling**. Why do I think that? Have you ever seen **Troop Beverly Hills**? If you haven't, we can't speak anymore because that is a great ass movie. **Jenny Lewis** , **Craig T Nelson** and **Shelley Long** were in that movie! IT'S ABOUT GIRL SCOUTS! IN BEVERLY HILLS! How could you NOT want to see it? I mean really? [... If **Jamie, Red Feather**, isn't enough proof then you have **Donna Martin**, the most boring annoying character in all West Beverly. Hey **Tori**, your Dad is the producer of one of the most popular tv show in the 90's and you going to play **Donna Martin**? How does that work? [...] I mean what did **Donna Martin** even do beside date **David Silver** and talk about not having sex. The only memorable thing I remember about **Donna Martin** was **that episode** she went to a dance dressed as a mermaid. I bet if I took a poll and asked which was your favorite **90210** character, **Donna Martin** or **Andrea Zuckerman**; **Andrea Zuckerman** would win hands down.

(ABOAGT/020708/ftsibaa)

The blogger clearly expects her readers to possess an extensive knowledge of classic US-American feature films and TV shows, featuring the actress *Tori Spelling*. The later (as well as the films in which she has played) is evidently expected to be shared knowledge between blogger and reader. Numerous characters of shows and films are simply mentioned without explaining their meaning or relation to each other. Hence the excerpt can only be made coherent with the knowledge of such necessary background information. For instance, one should already know that *90210* is the title of a popular TV show in the 1990's which revolves around the lives of a group of Californian teenagers. The proper names *Donna Martin*, *Andrea Zuckerman* and *David Silver* can only be properly integrated into a consistent [FILM] frame if one already knows that these are the names of the leading characters of the TV show *90210*. The fact that the blogger operates on an assumption of shared knowledge is made explicit when she declares that "if you haven't seen [the film] we can't speak anymore [...]"(ABOAGT/020708/ftsibaa). At a later stage, this blogger alludes to a

particular episode of the aforementioned TV show *90210*: “The only memorable thing I remember about Donna Martin was that episode she went to a dance dressed as a mermaid” (ABOAGT/020708/ftsibaa). This sentence can again only be made coherent if readers have watched the episode prior to reading the entry. Therefore, an extensive amount of specialised knowledge is necessary to fully comprehend this entry. The target readership of these entries is obviously one which is knowledgeable about US-American popular culture of the 1990’s. It is apparent that bloggers target their entries to users of a particular age by discussing films, music or TV shows of a specific era. The previous weblog entry in (84) focuses on users of the age between twenty and thirty years of age. They can be expected to be familiar with the TV shows, possibly watched in their childhood. Other entries written by other bloggers may target different age groups by discussing older TV shows or programs:

(85) I had to pick some old programmes [sic!] as I don’t ‘love’ much on TV anymore.

- \* Darling Buds of May (old)
- \* Only Fools & Horses (old)
- \* X Factor (and similar)
- \* Doctor Who

(BMNR/250708/nmydni)

(86) Entry: Here's an ear worm from the eighties that I had almost forgotten about.

Any time there is an Eighties Weekend on a radio station, this one will definitely be played. If someone mentions that they hate the music of the eighties, this song may be the reason why [...]

(BHA/180708/tsdmwh)

(87) Comment: Yeah, the 80's...when I listen to the "oldies" station on the radio I crank it up just to watch my daughters cringe! I consider it payback for listening to their music...

(BHA/180708/tsdmwh)

Example (85) illustrates how a different blogger lists her favorite TV shows in one of her weblog entries. From this list, the reader finds out about the blogger’s affinity to

“older” TV shows produced in the 1980’s (the TV show “Only Fools and Horses” was originally screened between 1981 and 1991 on British television). Further, one may infer from this list the blogger’s identity as a British national, because she enumerates only British rather than US-American TV shows. Hence, it is likely that the author primarily addresses a British readership rather than an American one. This hypothesis is sustained by the following exchange, which I found in another comment section of the same weblog:

(87) JD: The saddest part is that there was no donut. But that’s OK! We’ll have a belated celebration. CONGRATULATIONS!!!!!!

Babs: Donut?....donut? You’re on a British blog ‘ere you know!

(BMNR/230708/imtb)

Evidently, the user with the nickname *JD* spells the noun *donut* in the Standard American variety of English. The British blogger *Babs* reacts towards this fact by “reprimanding” the weblog user. She thereby implicitly conveys the cultural orientation of her weblog. In “correcting” *JD* and making explicit his linguistic deviation from British English, the blogger labels her weblog as a “British blog”. This move is likely to appeal to a British readership which shares with the author a similar cultural background.

Apart from that, this interesting excerpt illuminates the fact that the virtual identity of a blogger is construed not only by weblog’s topical or functional orientation but also by the way he makes use of a particular variety of language in his entries. If a blogger assumes a particular cultural orientation by choosing to write in a particular variety of language or by talking about certain topics (in particular ways), he likewise (re-)models his projected target readership. As blogger and user engage in the entry and comment section, both construct a collaborative discursive identity. They ultimately become partners in a *community of practice* (Eckert 2000:35) as co-participants in ongoing weblog discourse.

### 3. [TRAVEL]

The [TRAVEL] frame is the third most salient frame triggered by collocation relations in our weblog entries. This is not surprising given that bloggers habitually relate personal experience in their daily lives. The most reportable experiences in the lives of bloggers appear to be weekend trips, short and long excursions or vacations.

These events are usually narrated in great detail and complexity. The [TRAVEL] frame usually comprises a number of low level frames which specify the various procedural stages of a trip. Occasionally, bloggers report about traffic problems or other complications which may be related to the choice of a particular mode of transportation, e.g. problems to start, navigate or fit in a car, boat, plane, etc.. Apart from narrating about such difficulties, bloggers also describe the destinations of their travels, which is revealed by the following territorial collocations: *The United States - Europe, Canada - Europe, New York City - The Time Ball, America - Lady Liberty, Yukon Quest - Lance Mackay, Skagway - The White Pass McKinley - Alaska*. Some US-American bloggers seem to assume that their (US-American) users already have a substantial geographical knowledge about the regions and cities of the United States of America. Consider the following excerpt from a weblog entry:

- (88) Dad and I will be heading north in the morning to **the Mitten**. The 487<sup>th</sup> Bomb Group is having their annual reunion in **Kalamazoo** this year.  
(BHA/220708/tm)

Again, the blogger of this weblog entry presupposes that his readers can fill in the necessary background information needed to localise the city of Kalamazoo on the map of the United States. In addition, they are likely to be familiar with the meaning of the noun phrase *the Mitten*, which is addressed by a user in one of the following comments to the entry:

- (89) [...] I had to google 'Mitten,' Wikipedia took me to Michigan. Shaped like a mitten, dumb me. I learn some things some days. [...]  
(BHA/220708/tm)

Evidently, these assumptions of shared knowledge can be used to narrow down the projected target audience of a weblog. It is therefore a plausible assumption that the blogger of the previous weblog entry envisaged a US-American readership rather than an International group of addressees who might not be familiar with the geography of the United States. This applies, of course, even more crucially perhaps, to weblog descriptions of culture-related holidays and other cultural celebrations. Users need not only have an extensive knowledge of the historical background and

celebrations of certain cultures. Users also need specific procedural knowledge about the appropriate ways in which these cultural rituals are commonly celebrated. The following weblog entry was published on the fourth of July, the US-American Independence Day:

(90) HAPPY BIRTHDAY AMERICA!

[picture of the Statue of Liberty]

Luckily we tend overlook, or at least forget that Lady Liberty came from France!

She's just All-American to most of us!

[picture of the Statue of Liberty (different angle)]

Hope she is as proud of me as I am of her.

[picture of blogger in front of the Statue of Liberty]

HAPPY INDEPENDENCE DAY FROM ALL OF US!

(MM/040708/hba)

The insertions (picture one to three) in the entry stand for a number of photos of the Statue of Liberty in New York. In the entry, the statue acts as a collective symbol for the historical independence of the United States of America and its people commemorated the fourth of July. In this case, the statue and its symbolic force for US-American society is applied to trigger a collective act of memory. The Statue thus epitomizes a particular kind of cultural memory. In the US culture, the fourth of July has thus become what Assmann (1988:129) calls a "figure of memory", defined as a "fateful event in the past, whose memory is maintained through cultural formation [...] and institutional communication (recitation, practice, observance)".

With regard to these figures of memory, users naturally seek the symbolic and historic meanings of cultural memory induced by written or pictorial signs in weblog entries. Knowledge about the historical dimension of cultural holidays, for instance, comprises knowledge on the habitual rites which determine how the holiday is celebrated in certain cultures. For instance, it is customary in the United States to commemorate the Independence Day with a display of fireworks. Consider in this context, the following weblog entry:

- (91) Friday night, we went to my MIL's & FIL's house for a barbecue and fireworks. Yeah, I felt like crap, but feeling like crap and crapping is okay at their house because they are good taker-care-of-ers and, probably, I get better care there than I do at home. [...]  
(ShSe/090708/lme)

Apparently, the blogger here simply presupposes cultural knowledge on the customary celebrations of the Independence Day. Note that the holiday is never explicitly mentioned in the entry itself. One might suppose, on this basis, that the blogger primarily addresses US-American users or (at least) users who are familiar with the cultural practices of USA.

#### 4. [KINSHIP]

The frame of [KINSHIP] plays an important part in all AWC weblogs. Given the fact that personal weblogs characteristically deal with bloggers' personal life experiences, it is not surprising that they feature numerous instances of lexical items related to the [KINSHIP] frame. It is evident that bloggers are prone to repeatedly refer to close relatives in their weblogs whose identity needs to be introduced to weblog readers. Some bloggers invent nicknames for their friends and family to protect their real identities, others resort to generic names or kinship terms such as Mom, Dad, sister, etc. Relatives are usually introduced to users in one of the following four ways:

- (a) relatives or friends are introduced and characterized as part of the ongoing discourse in weblog entries.
- (b) relatives or friends are described in form of "friend lists" in weblog entries.
- (c) relatives or friends are presented through links to their Internet websites or weblogs.
- (d) relatives or friends are presented on separate web pages in the weblog.

The first possibility of introducing relatives and friends to weblog readers is to include them into the ongoing narrative of the entry. In the following weblog entry (92), the blogger presents the friend *Brooke* to her readership. As Brooke reappears in subsequent weblog entries, she is presented here as an important new friend of the blogger:

(92) On Monday, my new pal Brooke called me from Charles de Gaulle airport and asked me if I could log in to her email and check her mother's itinerary, because Mom didn't get off the plane. Of course! What are friends for? Mom took off on Monday June 30th from Los Angeles, but she didn't arrive until Tuesday July 1st in Paris. [...] So I met her at my nearest Metro stop and we had a nice leisurely walk to our favorite local, just one block from our apartment, Cafe Renaissance. We love this place for many reasons. [...] I waved cheerfully at Claude behind the bar. He came out to take our order and I introduced him to Brooke. It took him a while to understand her name. I told him that in English it's a name for une petite rive, which is wrong. Une petite rive means a small bank (of the river). I should have said un petit fleuve. He smiled anyway. He's not as strict, shall we say, as Zaina.  
(OMW/050708/htqs)

Note that the blogger characterizes Brooke indirectly by way of describing the reactions of the blogger's friends when they see Brooke for the first time. This is a subtle but very common way of an indirect characterization of recurring weblog characters. The second way of presenting relatives and friends is even more common in the AWC. This time, bloggers enumerate their acquaintances in linear order and characterize them briefly in few sentences. The following example excerpt illustrates part of such a list of friends which was posted in a single weblog entry:

Here's the list of this year's participants:

- Shoe of [SuperGurl](#): Shoe is the hostess with the most-ess. Shoe pulled off quite a coup getting this place. Boys Room sleeping in the bottom bunk below me. Who knew? I don't know how she does all this planning a
- That 1 Guy of [Drunken Wisdom](#): I have known T1G the longest of any of these guys. He's like a brother to n airport. We had a nice visit on the ride to the blogmeet. We shared the Boys Room and his steady snoring ro
- Erica of [Erica's Blog](#): Erica had quite the traveler's experience this year. She was four hours late getting into checked on the luggage when I picked up T1G, to no avail. Erica called several times from the house. She ha

Figure 31: Direct Characterization of Friends and Relatives in Weblog Entries

In figure 31, it is apparent that each individual characterization is accompanied by hyperlinks (green text elements). Upon activation, the hyperlinks lead users to the weblogs or websites of friends. User can thus use the new websites as springboard to finding out more about the various weblog personae. The usage of hyperlinks is thus another method to introduce and characterize people indirectly in a weblog.



**Ashley:** Oldest daughter. Used to be a pain in the ass most of the time; she's getting better.

**Awww-that's-so-cute:** Something that is so cute that it makes your eyes water.

**bliends:** Blog friends

**Chatterbox:** That bitch inside my head that tells me how worthless I am in every possible way.

**Cybersibling:** Someone you feel a connection to that is stronger than friendship. Often it is similar to that feeling you have for your biological siblings. For me it is [Mr. Fabulous](#), [Finn](#) and Da Pup.

**Dorkiness:** Actions that show how klutzy, weird, goofy, silly or stupid I am.

**Emily:** Middle child. Well behaved and intelligent. God must have known I needed a break after Ashley.

**FIL:** You guessed it, father-in-law.

**Gagalicious:** Something that is so bad, it makes you gag to think about it.

**Intertaining:** Strictly used to describe the cuteness of Isabelle.

**Isabelle:** Daughter of Ashley. Beautiful grandchild. Sunshine of my life. I love her dearly.

**Jamaicanese:** Jamaican language.

**Jason:** Wonderful husband of 16 years.

**Memers:** Bloggers who participate in the weekly memes.

**MIL:** Mother-in-law

**Motherly:** Acting like a Mom to someone who isn't your child, especially someone who is your peer.

**Reciprovisit:** Visiting someone because they visited you. Usually via weblog, but I suppose it

Figure 32: Presentation of Friends and Relatives on a Specific Web Page in the Weblog

As we can see in figure 32, some weblog authors even devote a particular page of their weblog to the characterizations of friends which are frequently mentioned in the weblog. Such a list can, for instance, be retrieved from the weblog *Shelli's Sentiments* in a section labeled "My dictionary" (cf. figure 32).

## 5. [INGESTION/FOOD], [PROFESSION], [COMMERCIAL], [TIME]

The remaining frames, which are frequently triggered by collocation relations in weblog entries, primarily relate to bloggers' culinary and commercial experiences both in their private lives and in their individual professions. The discursive revelation of the professions of bloggers may count as a further important marker of the blogger's identity. Some bloggers rarely write about their professional life at all. Other bloggers deliberately blend personal and professional experiences in an attempt to attract readers with similar professional backgrounds. The following weblog entry reveals a vast amount of facts about a blogger's work experience:

(93) When I was just shy of 20 years old, I got a job in pediatrics at the clinic where I am currently working. I worked for a female pediatrician for 7 years. I loved her (still do), she was like a Mom or an older sister to me [...] She has always been very good to me. The hard part about working for her is that she worked late hours and I had a young family that I needed to be there for. It just became too hard. For the next 6 years, I worked for several different **pediatricians**. [...] I feel comfortable working as a **pediatric nurse**. I just don't like working in **Family Practice** much. Ten years ago, if you had ask me if I would ever work **Urgent Care**, I would have said, "Hell no!" It terrifies me. It's too much like an **ER** and I don't feel confident at all with that type of intensity of care. They frequently send people via ambulance to the hospital from our **UC**. Partly because people often mistake our **clinic** for a **hospital** because it is that big.[...] People ask me all the time if I will ever go back to school for my **RN**. I always tell them the same thing.

(ShSe/220708/cz)

Weblog users who share a similar professional background with this blogger may have no difficulty in understanding the technical terms used in the above weblog entry. While some may still make sense to regular speakers of the English language (e.g. *family practice*, *emergency room*, *pediatricians*, *nurse*, etc.), other terms and differentiations clearly call for an advanced professional knowledge, e.g. that *UC* stands for *Urgent Care*, *RN* for *Registered Nurse*, or of the difference between a *clinic* and a *hospital*. The fact that the blogger does not explain these terms in her entry indicates that they are addressed to a target group of users with extensive knowledge about the medical professions and their related practical terminology.

The [TIME] frame is triggered whenever bloggers refer to temporal relations of hours, dates, months, years or seasons. The frequent use of time-related collocations underscores the exigency for a precise temporal localization of weblog entries within the temporal evolution of weblog writing. Usually, it is necessary for users to determine the exact time at which a weblog entry was written to make it coherent in its virtual context. In fact, users may sometimes need to browse through preceding entries in order to fully comprehend new weblog entries. Entries must therefore be sorted in chronological order so that users can locate and identify each related entry along a time line of an ongoing weblog discourse. The serial succession of entries

equally allows bloggers to “outsource” some important contextual information to their archive of past entries. Bloggers often resort to the practice of using hyperlinks in weblog entries to guide new users to background information accessible through previous posts:

(94) Happy Birthday, Baby!

Remember how much fun we had on my birthday last year? It was the big one.  
None of that fanfare this year. [...]

(ShSe/240708/hbb; underlined parts = hyperlinks)

(95) Proud Mommy

few weeks ago, I briefly mentioned the abscess that Sam had behind his right ear. After several days of antibiotics, it hadn't cleared, so the NP got us worked in with the ENT doc, Dr. K, at our clinic.

(ShSe/280708/pm, underscored segments = hyperlinks)

(96) Costa Rican Honeymoon Part 3

Part 3: 2nd Outing, Relaxing and Heading Home

See part 1 & part 2.

It's been 1 month since my last honeymoon recap and I think it's about time that I finished up, so I'll put the rest in this blog, and will try not to make it too long.

(FB/070708/crhp3, underscored segments = hyperlinks)

(97) And so Five Blondes began. As you can see, my initial idea had a different title. After a quick day of tossing around names, Micaela came up with the name “FiveBlondes.com”. [...] On July 17th, I blogged (warning: really boring). Then I introduced myself, followed by Micaela, Katelyn, Leah, and Lauren. Sorry for the lack of link for Leah, I can't find her intro.

(FB/16708/oyl, underscored segments = hyperlinks)

## 8.2 Serial Knowledge and Episodic Memory

In chapter three, I explained that there are two separate memory systems which human beings activate to make discourse coherent, i.e. *episodic and semantic memory* (cf. Tulving 1983, 2003). I defined semantic memory as the mental record of facts and procedures which are necessary for the engagement in everyday social life in a given culture or community. Episodic memory, on the other hand, is shared only by a limited group of individuals. It consists of “associations between items experienced at nearby times” (Howard et al 2005:2). In contrast to semantic memory, episodic memory therefore comprises an inherent serial nature. If two or more individuals share an episodic memory about a recent event, they share serial knowledge. Episodic knowledge may arise from the co-experience of the given situation or event by multiple individuals. In weblogs, it only seldom arises from the users’ actual co-experience in the realm of the ‘real world’. Episodic knowledge is nevertheless shared co-experientially, in the sense, that this happens on the discursively-mediated plane in weblog entries and comments. Over time, regular weblog readers come to share episodic knowledge as they are reported in weblog entries. New readers can acquire discursive episodes by following up previous entries and comments. I shall call this discursively generated subtype of episodic knowledge, *serial knowledge*. It is a hybrid category between episodic memory and what I have called *immediate knowledge* in chapter four. Serial knowledge is established through the narrative co-experience of reading weblog entries rather than the classic perceptive co-experience of an event in real life.

On this basis, we can now divide weblog users into three main types of addressees:

- 1) *regular user*: regular reader of a weblog who already acquired extensive knowledge about the weblog, its author(s), topics and the serial events posted in previous entries and comments
- 2) *frequent user*: frequent visitor of a weblog, only reads selective weblog entries interested in certain topics presented in the weblog (e.g. entries on hobbies, books, movies, music, professions, vacations, etc.)
- 3) *newbies*: new reader of a weblog who needs to familiarize him-/herself with the blogger, his or her virtual identity and topical interests, as well as the previous weblog discourse.

Bloggers usually write for a specific target audience, which usually consists of regular and frequent users. It seems apparent that the majority of regular users are individuals already known to the bloggers from their everyday lives, e.g. family, relatives and personal friends. Frequent users may be personal friends or online acquaintances. Newbies usually do not know the blogger personally and might have discovered the weblog on the Internet by chance. Figure 33 provides a final overview of the discursive interaction of collocation relations, shared knowledge, target readership and participation roles. The latter were developed in theory in chapter three and could now be applied to the empirical data in the AWC weblog entries.

We can see in the figure that there are two different (albeit interconnected) domains of the human memory, namely episodic and semantic memory. Episodic memory divides into knowledge acquired via co-experiencing a recent event in real-life (classic episodic knowledge) and shared knowledge about events revealed in weblog entries (serial knowledge). Users can attain a new participation status by acquiring serial knowledge in surfing and reading central entries of a weblog.

Semantic knowledge can be differentiated into communicative knowledge about ordinary, culture-bound events and procedures, cultural knowledge about ritualistic acts of celebration and commemoration across cultures and linguistic knowledge. Collocation relations in weblog entries reflect the most salient cognitive frames which induced by bloggers in weblog entries. The frames provide knowledge fact-related and procedural knowledge about the various topical (and sub-topical) domains of discourse. Through a hybrid interaction of discursive engagement in comments and cognitive engagement in entries, users can claim for themselves certain participation roles by

- making themselves known to the blogger. They can recurrently post comments on the weblog to seek the attention of the blogger or they can alternatively use other communication channels to achieve this effect, e.g. telephone, email, etc..
- By acquiring or sharing similar kinds of episodic, serial, linguistic or encyclopaedic knowledge with the blogger. As a result, they are usually addressed as the target audience of the weblog and peruse the weblog as acknowledged “insider”.

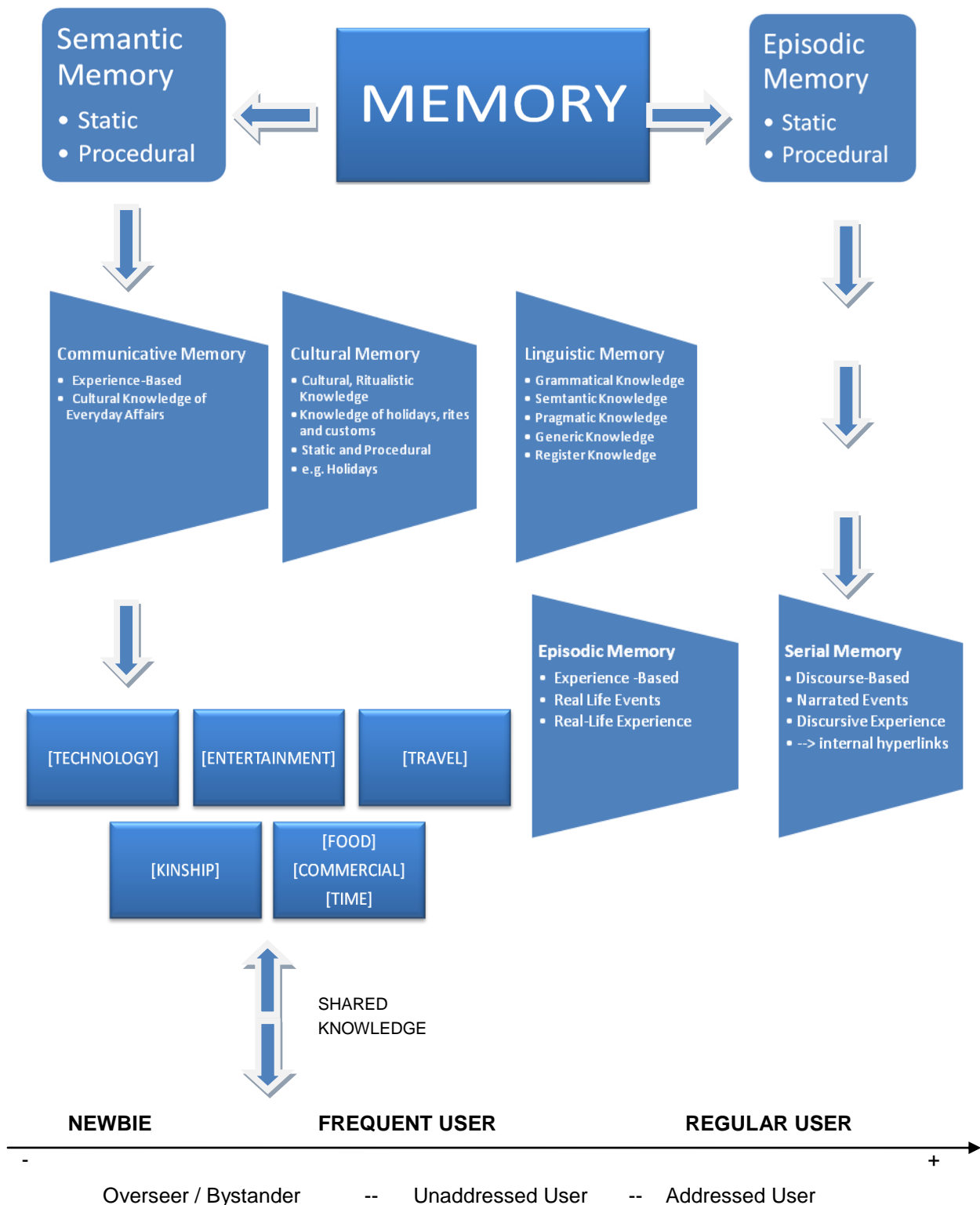


Figure 33: Collocation, Participation and Knowledge in Weblog Entries

As shown previously, the target readership of weblogs may be deduced from (a) the type and (b) the frequency of the (topical) frames which underlie and structure collocation relations in weblog entries. The frequency of individual frames is

indicative of the concrete topical orientation of each weblog. The frame types specify the domains of knowledge, which have to be activated when users try to turn weblog entries into coherent, unified wholes. In order to understand and engage in the discourse within a particular frame in and across weblog entries and comments, users must share the same frame knowledge with the weblog author. In the AWC weblog frame knowledge was found to pertain to the following fields of memory based on a topical classification of collocation relations:

- factual and procedural knowledge about various institutions of a culture (e.g. schools, hospitals, universities, supermarkets, etc.),
- geographical knowledge about places, federal states or countries (e.g. Mitten = Michigan, Kalamazoo, etc.),
- terminological knowledge about specific areas of interest, hobbies or professions (e.g. medical lingo, films, books, music, etc.),
- linguistic knowledge about registers and varieties of a given language (e.g. donut vs. doughnut),
- cultural knowledge about collective acts of remembrance in a given culture (e.g. holidays, customs, etc.),
- serial knowledge of previous weblog entries,
- episodic knowledge of previous shared real-life experiences.

An examination of the terminology most commonly used by bloggers helps to narrow down the main target readerships. Accordingly, their areas of interests, and their extensive knowledge thereof, typically revolve around the following domains:

- English-speaking,
- familiar with US-American (of US weblogs) or British (of British weblog) customs and culture respectively,
- knowledgeable in different distinct areas of interest (e.g. music of the 1980's (BHA), contemporary pop music (OMW), US-American TV shows of the 1990's (BOGT), farming and agriculture (BHA), parenting (AA), medical life (ShSe), etc..),
- familiar with the blogger, his or her relatives and friends,
- well versed with all or most of the weblog entries.

We can conclude that weblog users can either be addressed explicitly in weblog entries and thus assume the role of addressed users or remain unaddressed but implicitly acknowledged by the author. Both addressed and unaddressed users comprise the target readership of weblogs as long as their identity is known to the weblog author. Both addressed and unaddressed users usually have extensive knowledge about the various fields listed above. Both are representatives of the user group, i.e. *regular users*.

Unaddressed users, who are not known to the blogger but who frequently read specific weblog posts, are called *frequent users*. They may share knowledge in particular areas of interest but not in others. Newbies, who have just started reading a weblog and who are usually unknown to the weblog author, are *overseers / bystanders*. In actively participating in the comment sections or following the weblog over a considerable period of time, they can turn into frequent users, or even new regular users of the weblog. An accumulation of weblog-specific knowledge thus promotes newbies to full membership in a weblog community. Such a blogging community is defined by a mutual alignment of common ground (shared knowledge) between a blogger and his (or her) users as well as their discursive alignment of a particular use of language between the entries and comments. I have shown that one way to retrace this discursive alignment is through the analysis of collocation in weblogs.

### **8.3 A Brief Analysis of Weblog Interaction**

Finally, I shall answer the question to which extent one can speak of a dialogical interaction between weblog entries and comments. It was stressed throughout this study that the interaction between weblog entries and comments is severely limited compared to the interactive complexity of two-party or three-party conversations. I have so far evaluated the cohesive interaction of weblog discourse and will devote my attention now briefly to a final look at the speech acts involved in weblogs.

Weblog, it was demonstrated, possess a monologic structure. Responsive or reactive speech acts are concurrently scarce. In fact, they barely exist in weblog entries. The only responsive speech acts found in a weblog entry were elicited in the collaborative weblog *Five Blondes* and this may be due to the collaborative design of this weblog which clearly distinguishes it from others:



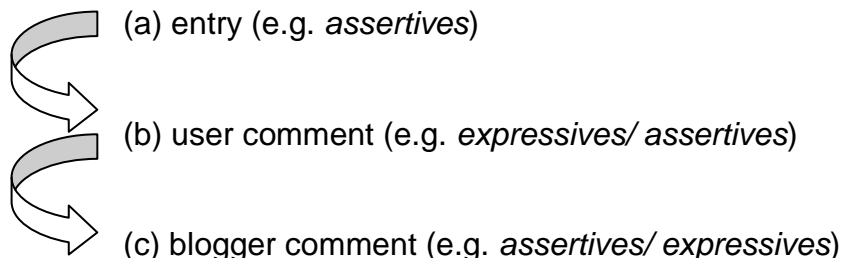
(98) Picture Every Day for a Year

Thanks for all the great suggestions! I've decided that I'm going to do a summer theme for the first month, starting tonight at a barbeque!

I'll post my flickr account later tonight!

(FB/060708/pedfay)

The expressive speech act “Thanks for all the great suggestions!” does not reply to a specific comment made in reference to a previous entry but can be interpreted as an expression of gratitude toward a number of previous comments. The concept of addressing multiple rather than individual users at once underscores the fact that one-to-one interaction is extremely infrequent in weblog entries. In weblog comments, one-to-one interaction between different weblog users is equally rare. The most common type of interaction in weblog comments is the one-to-one interaction between blogger (in blogger comments) and user (in user comments). Blogger comments are often integral parts of a recurrent exchange and speech act pattern which prototypically follow three cumulative stages:



The following is a prototypical example of such an exchange pattern:

(99) Julie Gong (entry): [...] The whole time I've been listening to the [music] album at work I thought it was in English. It clearly isn't. The worst part might be that I tried to argue with him [a friend] about Gobbledigook. I told him my version was in English. Wrong. [...]

Stephanie (comment): It's okay that you didn't know [that the album was not in English]. I have similar moments.

Julie Gong (comment): stephanie - thank. thank you very much.  
(ABOAGT/080708/srini)

We can see that, initially, the entry establishes a topic and sets the emotive tone for the following exchange. In the second stage, a user responds to the views presented in stage one. This is commonly done in an acknowledging and appreciative manner. Finally, the blogger herself responds to this user comment inside the comment section, thus providing closure to the exchange.

There are three interesting facts to be gained from the above example: first, the length of individual contributions recedes from entry (here: 528 words) to the user comment (here: 17 words) to the final blogger comment (here: 6 words). A decline of word length has been found to be a salient feature in these kinds of weblog exchanges. This feature is often indicative of a simultaneous decrease in reference relations because long texts usually host more substantial numbers of personal and demonstrative reference than shorter ones.

A second interesting fact is the frequent use of ellipses in the exchanges. In the above example, for instance, we can immediately identify three overt uses of verbal and clausal ellipsis both within and across individual contributions:

- (1) It clearly isn't [sung in English] → *verbal ellipsis*
- (2) [I was] wrong → *verbal ellipsis*
- (3) It's okay that you didn't know [that the album was not in English] →  
*clausal ellipsis*

The frequent use of ellipsis can be explained on textual and interpersonal grounds. On a textual plane, the implicit transfer of a particular part of information from one contribution to the next indicates shared knowledge between blogger and user. On an interpersonal plane, we may argue that ellipsis across contributions trigger a certain degree of interpersonal alignment because a part of one's own discourse has been appropriated by another. The interactive overlap of elliptical constructions builds up interpersonal relations between the interlocutors.

The third fact relates to the types of speech acts which are prototypically involved in the exchange pattern. Entries predominantly comprise assertive speech acts, while expressive come in very low numbers. The assertive is the speech act mode which bloggers adopt when they present, describe or narrate some personal

experience or feeling. Here are some excerpts excerpted from different weblog entries:

(100) I went on a date last night (*assertive*) and I was hoping something ridiculous would have happened (*assertive*) because ridiculous thing usually happen (*assertive*) when I'm around this person (*assertive*) but you are about to be disappointed (*assertive*). Not one remotely funny thing happened (*assertive*)  
(ABOAGT/110708/jn)

(101) Before I launch into writing (*assertive*), I want to tell everyone that I am changing D's blog name (*assertive*). He will no longer be known as 'D' (*assertive*) so that I no longer have to begin sentences nonsensical (*assertive*) explaining what "D and I" did (*assertive*).  
(FB/090708/dtpaoswd)

(102) With kids, long weekends aren't something to be looked forward to as much as they used to be (*assertive*). Actually, that isn't fair to AE (*assertive*). With infants, long weekends aren't something to be looked forward to (*assertive*). It is just really hard to find something to do with a 6-month-old all day! (*assertive*)  
(NSM/070708/aws)

Responding to these entries are user comments. They mainly consist of expressive speech acts in which users express their (usually positive) evaluation toward one of the focal topics or elements introduced in the related entry. Many users begin their comments with an expressive speech act and then modify it with an additional assertive presentation of their own. For instance, they may first state that they like the point of view expressed by the blogger and then give reasons for their positive evaluation. Alternatively, they may add their own personal impressions or past experiences. Here are some speech act analyses of user comments from different weblogs, all of which evidently comply with the aforementioned pattern:

(103) Terrific recap (*expressive*). You captured it perfectly (*assertive*)  
(BHA/010708/tb2008)

- (104) Thanks again for taking me to the airport Jerry (*expressive*) I'll bet you and Nora were tired after all that driving! (*assertive*)  
(BHA/070708/bfn)
- (105) Wheeee! Thanks for driving old man (*expressive*). Work was rough today (*assertive*), but I'm feeling much better after a big old nap (*assertive*).  
(BHA/070708/bfn)
- (106) very cool! 😊 (*expressive*) I tried new and snazzier template (*assertive*), but I gave up (*assertive*) ... I'm sticking with good ole blogger (*assertive*).  
(BMNR/010708/nbtoii)
- (107) Babs - I adore your theme (*expressive*). It's so welcoming and friendly (*assertive*), just like it's owner! (*assertive*) You may know how insane templates make me (*assertive*) [...]  
(BMNR/010708/nbtoii)

Most user comments followed this prototypical speech act sequence. Since all comments are presented in linear order on screen, and comments are written and published asynchronously, there are naturally no overlaps between the contributions of different speakers. *Turns* are not allocated or negotiated between users and bloggers. The boundary of turns in weblog interaction therefore amounts to the perceptual limits of comments on screen, including their underlying character limitations. Turn length and turn position are partly constrained by the technological setting of weblog discourse.

Finally, the organisation of speech acts in blogger comments generally copied the speech act profile described for user comments. Nonetheless, expressive speech acts seemed to prevail over assertive ones.

Even weblogs which consisted of quite extensive lists of comments did not feature longer or more complex interaction patterns. It would be theoretically possible for users to reply to blogger comments. However, in the AWC corpus no such reply could be identified. Concluding, I propose five central tenets which delineate the potential and boundary of weblog interaction:

1. The types of speech acts involved in weblog interactions are extremely limited. Interlocutors mainly resort to assertive and expressive speech acts, which mostly follow a strict sequential pattern.
2. The interaction is always initiated and guided by weblog authors. Weblog entries do not usually finish with first-pair parts like questions. Rather, the invitation to respond to weblog entries seems already presupposed by the general purpose of the blogging practice.
3. Interaction in comments is also extremely limited. The exchange pattern is mostly advanced by weblog authors in their responses to user comments. Users do not reply to blogger comments although this would be both technologically and discursively viable.
4. The length of contributions in weblog interactions decreases greatly from the initial stages of weblog contribution in entries by bloggers to their own response-responses in blogger comments.
5. Weblogs comments regularly feature nominal and verbal ellipsis between user and blogger comments, which partly explain the decreasing length of contributions from one response to the next. They are equally indicative of the interpersonal alignment between both interlocutors.

Alongside, the findings reported and interpreted in the analysis of grammatical and lexical cohesion, we can contend that the degree of speech act interaction between weblog entries and comments was found to be largely constrained in the empirical data analysed. The constraints reinforce the aforementioned claim that both weblog entries and comments primarily subscribe to a cohesive profile which is much closer to written monologues than to spoken dialogues.

## The Results

---

### 9.1 Cohesive Interaction Revisited

The various results observed and discussed in this study finally need to be reconciled with the research objectives devised in the beginning of this thesis. I started this investigation with the assumption that the discursive realm of weblogs is still unexplored by linguists. The aim of this study was therefore twofold:

- (a) to explore the cohesive and discursive interaction within and across weblogs,
- (b) to explore the extent to which the cohesive texture of weblogs is similar to written monologues and spoken dialogues.

In order to determine the communicative status of personal weblogs, I developed a suitable framework for the quantitative and qualitative analysis of weblog cohesion.

At the same time, a corpus was devised, which enabled me to apply the cohesive framework to authentic weblog data from entry and comment sections. I could now contrast the distribution and types of cohesive means in weblog entries and comments with the cohesive distributions of other text genres. For the purpose of this study, academic articles and conversations were chosen as the most characteristic text genres for the written monologues and spoken dialogues respectively. The quantitative distributions of weblog entries, comments, academic articles and face-to-face conversations were mapped onto each and discussed by recourse to a range of qualitative weblogs examples. This methodology allowed a multi-dimensional view of the monologic or dialogic use of various grammatical and lexical means of cohesion. The individual results of this analysis could therefore be used to draw a comprehensive picture of the monologic or dialogic texture of weblog cohesion.

This method of analysis only focused on the quantitative and qualitative analysis of cohesive means. It excluded all non-cohesive linguistic features from its analysis, which have often been included in other typological studies on weblogs (cf. Puschmann: forthcoming (b)). Rather, it was argued that the cohesive analysis chosen in this study captures more comprehensively the discursive realm of weblog which is responsible for determining the monologic or dialogic character of weblog discourse.

At the beginning of this study, I postulated four basic research questions which guided both the theoretical and methodological approach of this study. It seems sensible now at the very end of this thesis to give some adequate concluding answers to each one of them:

- Do weblog entries and comments make different use of cohesive means?
- How exactly do weblog entries and comments differ with respect to their variation and distribution of grammatical means of cohesion?
- How exactly do weblog entries and comments differ with respect to their variation and distribution of lexical means of cohesion?
- Does the cohesive profile of weblog discourse resemble the spoken dialogue, the written monologue or both?
- How interactive is weblog discourse in entries and comments?

## **9.2 Monologue or Dialogue? – Positioning Weblog Discourse**

I shall answer the previous questions in consecutive order, starting with the first one:

- Do weblog entries and comments make different use of cohesive means?

One surprising result of this study is the fact that the AWC weblogs possess very similar cohesive profiles. The close resemblance of weblog entries and comments can be interpreted as the result of the communicative conditions under which both types of discourse are composed. Hence the entries and the comments mainly include written language. Furthermore, the asynchronous interaction between weblog entries and comments frees bloggers and users from the productive and procedural constraints of spoken interaction, i.e. *turn-taking* etc. *Overlaps* and *simultaneous speech* are absent from the weblogs since the weblog software automatically aligns all contributions in reverse chronological order on the screen.

Furthermore, spontaneous or erratic contributions which have been shown to be common to Internet relay chats (cf. Beißwenger 2005) are infrequent in the AWC. All entries and most comments contain long stretches of discourse (more than seven sentences) which exhibit a high clausal complexity as well as lexical variety. These observations all point to an extensive planning time involved in production and processing of weblog entries and comments. In addition, weblog users can

deliberately access and re-read passages at various stages of their immersion in weblog discourse. They do not have to cope with the medial fluidity which governs the discourse of spoken language. In the conversations, for instance, interlocutors have to compensate for the perceptive transience of their voice by using a wide array of repetitions. It follows that similar communicative conditions yield similar cohesive profiles of the entries and comments. This brings me to the next research question:

- How do weblog entries and comments differ with respect to their variation and distribution of grammatical means of cohesion?

All categories for grammatical and lexical cohesion introduced in chapter four have been found in the AWC corpus. Some categories were particularly frequent in both weblog entries and comments (*personal and demonstrative reference, ellipsis*), others were only found in small quantities (*comparative reference, substitution, conjunction*). The following cohesive pattern emerged from the analysis of grammatical cohesion: the most frequent category of grammatical cohesion in weblog entries and comments was personal and demonstrative reference, followed by comparative reference, ellipsis, substitution and conjunction. Personal and demonstrative reference in comments occurred in three different varieties: (a) reference within single comments (*internal reference*), (b) reference across comments (*cross-comment*), (c) reference between comment and related entry (*entry-related*). Internal reference was most frequent in comments, followed by entry-related reference. Only few instances of cross-comment relations could be elicited in the data. Hence, cohesive interaction between comments could hardly be documented at all. It follows from these observations that the majority of weblog comments in the AWC are cohesively autonomous, self-contained units of discourse.

Entry-related reference proved to be particularly frequent in the comments which responded to visual or audio stimuli embedded in related entries, e.g. a video, photo or audio file. Intertextual reference between comments and entries therefore regularly includes pictorial elements which function as cohesive anchors for both weblog text units, entries and comments.

*Comparative reference, substitution and conjunction* were much less frequent than personal and demonstrative reference in the weblog entries and comments. No cross-comment or entry-related relations of these three categories could be elicited in



the analysis. All instances were internal relations, relating discourse elements within single weblog comments. Interestingly, substitution was not found to be more frequent in weblog comments than in entries. The analysis of conjunction across the entries and comments proved that the weblog entries exhibit more additive and adversative conjunctions than weblog comments. In contrast, the weblog comments, show higher ratios for causal and temporal conjunctions. This distribution of conjunction types indicates that weblog entries feature a consistent narrative dimension, whereas comments may be either narrative or argumentative in kind.

*Ellipses* were more frequently used in weblog comments than in entries. In fact, a few cross-comment and entry-related relations of ellipsis could be identified in the comment sections. The frequent use of ellipsis in the weblog comments is reminiscent of spoken registers (cf. *netspeak*, Crystal 2001). I argue that the frequent use of ellipsis and interpersonal pronouns imitates dialogic discourse in the weblog comments. It is possible that this personalised style of writing is largely responsible for the ongoing myth of dialogic interaction in weblogs.

Comparatively low frequencies for cross-comment and entry-related reference indicate that cohesion in weblog comments is primarily sustained within single text units. The dominance of same-speaker relations is also characteristic of written monologues rather than spoken dialogues. A comparison of weblog frequencies for personal and demonstrative reference to results of a previous study on pronominal reference across text genres (Biber 1992) confirmed the hypothesis that the cohesive texture of weblog entries and comments is mainly of monologic nature.

- How do weblog entries and comments differ with respect to their variation and distribution of lexical cohesion?

The monologic tendency of weblog entries and comments which surfaced during the analysis of grammatical cohesion could be largely sustained by the analysis of lexical cohesion. Both weblog entries and comments showed high frequencies for total and partial recurrence as well as for elaborative collocation. The high number of collocation pairs in the weblogs moves both the entries and the comments closer to the academic articles as these typically possess more collocation pairs than the two-party or three-party conversations. Weblog entries and comments equally have lower frequencies for total recurrence than the conversations although their mean scores

for this category are higher than in academic articles. The high frequencies for total recurrence in weblog entries and comments can be explained by their frequent use of interpersonal pronouns. Once I deleted the pronouns from the list of total recurrence relations, the entries as well as the comments exhibited a striking similarity to the profile of total recurrence relations in the academic articles. Although the latter frequently include more antonyms and co-hyponyms than the entries or comments, the quantitative difference is largely insignificant.

On the whole, I can conclude that the cohesive distribution of grammatical and lexical cohesion in weblog entries and comments is similar. Their cohesive profiles are much closer to the cohesive distribution of the written monologue (academic articles) than the spoken dialogue (two-party conversations). To this end, let us finally take a detailed look at the interrelation between the spoken dialogue, the weblog and the written monologue.

- Does the cohesive profile of weblog discourse resemble prototypical spoken dialogues, prototypical written monologues or both?

In the course of this study, I have developed a methodology for the compilation of language data in weblogs. This resulted in a weblog-specific framework of cohesive relations which I applied to a self-compiled weblog corpus including entry and comment sections. A number of viable quantitative and qualitative results emerged from this application, and they have been presented and discussed exhaustively in the previous two chapters. The cohesive profiles of weblog entries and comments were compared to a recent study on the cohesive variation of spoken and written text genres (Tanskanen 2006). In comparing the cohesive distribution of weblog entries and comments to the prototypical profiles of academic articles and conversations, a specific degree of similarity between the text genres could be ascertained. Finally, I can now reconcile the different parts of the analysis and determine the actual discursive status of weblog discourse as being monologic, dialogic or hybrid. The following table summarises the cohesive profile of weblog discourse (in entries and comments) in contrast to written monologues and spoken dialogues:

	Weblog Discourse	Written Monologue	Spoken Dialogue
<b>Pronouns</b>	medium anaphoric low exophoric	medium anaphoric low exophoric	high anaphoric high exophoric

	low cross-comment low entry-related low vague pronouns	no vague pronouns	high vague pronouns
<b>Interpersonal Pronouns</b>	high frequency	low frequency	high frequency
<b>Substitution</b>	low frequencies	low frequencies	high frequencies
<b>Ellipsis</b>	<i>Entries:</i> low frequencies	low frequencies	high frequencies
	<i>Comments:</i> medium - high frequencies		
<b>Conjunction</b>	<i>Entries:</i> high additive / adversative <i>Comments:</i> high temporal / causal	<i>Fiction:</i> high additive / adversative <i>Religious:</i> high causal <i>Journalism:</i> high temporal	inconclusive data
<b>Recurrence (+IPPs)</b>	medium total recurrence	low total recurrence	high total recurrence
<b>Recurrence (- IPPs)</b>	low total recurrence	low total recurrence	medium total recurrence
<b>Collocation</b>	high frequencies	medium frequencies	low frequencies
<b>Co-Hyponymy</b>	low frequencies	medium frequencies	medium frequencies
<b>Other lexical categories</b>	low frequencies	low frequencies	low frequencies

Table 63: Weblog Discourse (Entries and Comments) vs. Written Monologue and Spoken Dialogue  
(■ = Dialogue Profile, ■ = Monologue Profile, ■ = Intermediate Profile (closer to monologue))

In table 63, it is evident that weblogs and written monologues share more cohesive similarities than weblogs and spoken dialogues. Weblogs share the frequent use of interpersonal pronouns and the frequent use of ellipses in comments with spoken dialogues. This similarity bears witness to an informal and abridged style of writing in some weblog comments which simulates a conversational style of writing in weblogs. Nonetheless, a closer look at the cohesive texture of weblog discourse reveals that the interaction within and between entries and comments is extremely limited. Most entries and comments are individual, autonomous monologues which showed only a few intertextual bonds. The low frequency for cross-comment relations and the fairly

restricted use of entry-related reference in comments bears witness to this fact. Another proof is the low frequency for substitution in entries and comments and the low frequency of ellipsis in entries. More evidence was provided by the high collocation frequency in weblogs which showed more similarity to written monologues than to spoken dialogues. Additionally, the frequency for total recurrence in weblog discourse proved to be more analogous to written monologues than to spoken dialogues. Once the interpersonal pronouns are factored out of the equation, weblog discourse showed a striking similarity to written monologues. Generally speaking, I conclude that weblog discourse shows multiple cohesive analogies to written monologues while it shows some crucial differences to spoken dialogues. As a result, I consider personal weblogs as written monologues, which exhibit only few features of prototypical spoken dialogues. The frequent use of interpersonal pronouns and ellipsis are the only two viable indicators of the spoken dimension in weblogs. It seems questionable to classify personal weblogs as hybrid forms of discourse based only on these two empirical findings. We have more evidence for the classification of personal weblogs as text genres which are very similar to written monologues. This classification leads me to the final research question:




- How interactive is weblog discourse in entries and comments?

### **9.3 Communicative Conditions in Personal Weblogs**

The previous analysis of grammatical and lexical cohesion in weblog discourse has shown that the cohesive texture of weblog discourse and written monologues are in many respects analogous. It has already been stressed on many occasions that both types of discourse emerge under similar socio-communicative conditions. Table 64 and 65 depict the most salient discourse-related and media-related characteristics of weblogs (entries and comments), written monologues and spoken dialogue. Both tables therefore capture the most elementary similarities and differences across these three text genres. The various “facets” used to describe the technological preconditions for Internet discourse have been selectively adopted from a classification scheme for computer-mediated discourse proposed by Susan Herring (2007)

General Features	Personal Weblogs	Written Monologue	Spoken Dialogue
Salient Mode of Production	Writing	Writing	Speech
Participation Structure	One-to-Many Many-to-One	One-to-Many	One-to-One
Text/Discourse Type	Narrative Argumentative	Argumentative	Conversational
Genre	Public Journal	Academic article	Casual Chat

Table 64: General Discourse Features in Three Text Genres

( = Dialogue Profile,  = Monologue Profile,  = Intermediate Profile (closer to monologue))

Medial Features	Personal Weblogs	Written Monologue	Spoken Dialogue
Synchronicity	Asynchronous	Asynchronous	Synchronous
Message Transmission	Message-by-Message	Message-by-Message	Character-by-Character
Persistence of Transcript	Medium – High	High	Low
Size of Message Buffer	Constrained by Weblog Software	Constrained by Document Size	Unlimited
Channels of Communication	Visual Modes	Visual Modes	Auditory Modes

Table 65: Medial Characteristics in Three Text Genres

( = Dialogue Profile,  = Monologue Profile,  = Intermediate Profile (closer to monologue))

Table 64 and 65 clearly demonstrate that weblog discourse and written monologues share similar productive and receptive conditions while they both are very different from spoken dialogues:

- Both weblogs and written monologues comprise visual means of communication (written language, pictures, etc.). They equally apply written language as their primary mode of production.
- The most common participation structure in weblogs and written monologues is *one-to-many* which implies that weblog authors mainly establish and continue interactions in weblog entries and comments. Weblog users seldom interact with each other in comment sections. Cohesive interaction may include a limited amount of *many-to-one* interactions between users and weblog author. However, I have shown in the previous chapter that the range of speech acts used in these interactions is as limited as their exchange

pattern. Interactions are predominantly triggered and entertained by weblog authors. Consequently, weblog users only partially embrace the potential for two-way communication in weblogs. Most contribute fairly autonomous monologic comments to the weblog which comprise some lexical but remarkably few grammatical ties with related entries or comments. Both cohesive and discursive interaction in weblogs is extremely constrained.

- Weblog entries and comments have been shown to feature narrative and argumentative text types which can equally be found in written monologues (e.g. fiction and scientific articles) as well.
- Both personal weblogs and written monologues share an asynchronous discourse setting which enables their interlocutors to plan their contributions before the discursive engagement or publication of the texts. Also, readers and users can review contributions over a longer period of time which facilitates the comprehension of longer stretches of discourse. The persistence of transcript in both types of discourse is high while spoken dialogues are much more fluid in this respect.<sup>82</sup> The transience of discourse has been shown to influence the cohesive strategies used in the different text genres. I suggest that it is one of the main reasons for the cohesive variation discussed across weblogs, written monologues and spoken dialogues in the previous two chapters.
- According to Herring (2007), the feature *size of message buffer* indicates the “the number of characters the system allows in a single message”. While in weblog entries and comments the number of characters allowed for each text section is pre-determined by the specific weblog service used (e.g. wordpress, blogger, etc.), the number of characters in academic articles is constrained by the document size on which they are printed (A4, A5, A7, etc.). In this context, spoken dialogues can principally include unlimited numbers of characters. The limits of spoken interaction are not determined by material constraints but rather by social conventions.

---

<sup>82</sup> Clearly, one has to take into account here the relative fluidity of weblog discourse. Some weblog entries and comments may be deleted by bloggers and this can seriously impede user comprehension of certain weblog parts. However, such alterations or deletions of text units from weblog archives are still exceedingly rare.

Note that both topic and purpose of personal weblogs and academic articles (written monologues) are rather different. On the one hand, academic articles are extremely constrained in form (convention, style, composition, text design) and topic (academic themes). They seek an objective and rational tone and rigorously exclude any references to personal evaluations. On the other hand, personal weblogs embrace and encourage a personal and evaluative style of writing. They revolve around everyday affairs and include long narrative stretches of discourse. One could expect these differences in discourse style, topic and purpose to surface more clearly in the cohesive profiles of both text genres. However, a considerable difference could not be elicited in the AWC data. Rather, both weblog genres were found to be quite similar to each other with respect to their use and distribution of cohesive means. What remains is the correlation between similar communicative conditions of weblogs and written monologues and their resemblance of cohesive profiles. One is tempted to interpret the communicative conditions as the determining factor for the similitude of cohesion in both text genres. However, it is neither the technology nor the software of weblogs which creates discourse on weblogs. Rather, the cohesive texture elicited in the previous chapters reflects the way bloggers and users interpret their role and participation in weblog discourse. Although weblog software technically allows for more participatory exchange patterns in weblog comments, weblog users mainly content themselves with writing single responses to weblog entries. Although the socio-technological conditions of weblog communication surely have an impact on the way bloggers and users engage in discourse, they cannot be made solely responsible for the monologic undercurrent of personal weblogs.

On this basis, I conclude that the weblog interaction can only be explained by comparing the discursive (cohesive) profile of texts to the technological exigencies imposed by specific media (and forms of communication). Weblog discourse is precisely what happens when both meet and converge to form a new writing space in the online sphere.

#### **9.4 Limitations of the Study and Future Research**

Any scientific study necessarily has to account for its theoretical and methodological limitations. This one is no exception. As a result, a few critical remarks should be voiced at the very end of this study, not least to encourage further research in this promising and intriguing area of linguistic study.

The study itself only considered a limited number of personal weblogs. This quantitative restriction was necessary due to the time-consuming nature of the manual analysis involved in eliciting cohesive relations in and across weblog entries and comments. Clearly, this methodological choice entails ramifications with respect to the representativeness of language data for the genre of personal weblogs. An inclusion of additional weblogs would have possibly produced different results than the ones discussed in this book. Nonetheless, I argue that a further extension of the data sets in the AWC would have only been viable if multiple analysts had engaged in the analysis of this study simultaneously. As this study was designed to be conducted by only one person, this option was already excluded at a very early stage of planning. The incorporation of further language data into the AWC would only have been possible at the expense of close qualitative analyses of the language data. Finally, the notion of representativeness itself can be challenged. So, Kennedy (1998:62), for instance, remarks that “notions of representativeness and balance are, of course, in the final analysis, matters of judgment and can only be approximate”.

Biber (1988:53) reveals another classic restriction of corpus-related work: the assignment of “undue weight to individual texts.” He continues that “unusual or idiosyncratic texts can have a major influence on the analysis” (Biber 1988:53). Therefore, weblog data chosen for this study were carefully assessed in regard to their generic and formal and socio-linguistic compatibility. They were analysed with the help of functional (questionnaires), topical (tag cloud analysis) and formal criteria. As a result, I could guarantee that only a generically-compatible number of weblogs entered the AWC data.

Moreover, weblogs were evaluated only with respect to prototypical text genres representing either spoken dialogues (two-party and three-party conversations) or written monologues (academic articles). The additional comparison of written dialogues (letters, message boards) or spoken monologues (speeches) would have been informative and invite more future research on the topic. Another intriguing point of departure for linguistic studies is the multimodal analysis of cohesive relations between weblog texts and pictures. As indicated before, reference relations between weblog entries and comments largely include the cohesive use of pictorial elements. The cohesive contribution of pictures and videos in weblogs definitely deserve a close investigation.



## 9.5 Concluding Remarks

The principal goal of this study was to investigate the degree to which interlocutors collaborate to construe cohesion in and across weblog entries and comments. I could reveal that, despite their wide-spread reputation as “written dialogues”, weblog discourse actually exhibits a low degree of cohesive interaction. Consequently, the cohesive profile of weblog discourse was found to be closer to written monologues than to spoken dialogues. In line with this finding, the form and function of speech act patterns between weblog entries and comments was highly constrained. Rather than being interactive on a discursive plane, weblog authors and users are highly (inter-)active on a cognitive, hermeneutic plane. For instance, the analysis of collocation relations *inter alia* revealed that bloggers and users are actively involved in a range of cognitive activities which allow them to make sense of specific weblog entries (or comments). These activities may include the acquisition of serial knowledge acquired through reading previous weblog entries or the establishment of communicative or cultural knowledge needed to adequately frame the event or affair discussed in the weblog. In contrast to the actual verbal interaction in personal weblogs, this more subtle cognitive (inter-)action between weblog author(s) and user(s) is very much alive. In other words, the monologic nature of personal weblogs shifts relocates interaction to a cognitive plane. The key to a better understanding of weblog cohesion and discourse is likely to be found at the crossroads of discursive and cognitive interaction.

## References and Bibliography

- 
- Altenberg, Bengt.** 1986. Contrastive linking in spoken and written English. In Gunnel and Ingegerd Bäcklund Tottie (ed.). *English in speech and writing: a symposium*, 13–40. Stockholm: Almqvist and Wiksell.
- Ammon Ulrich et al** (ed.). 2004. *Sociolinguistics. An international handbook of the science of language and society*. Berlin: de Gruyter.
- Androutsopoulos, Jannidis.** 2008. Potentials and limitations of discourse-centred online ethnography. *Language@Internet*(5). No pagination (online publication). <http://www.languageatinternet.de/articles/2008/1610> (accessed December 30th 2009).
- Ansel Suter, Bettina.** 1995. *Hyperlinguistics. HypertextLernumgebungen im akademischen Kontext: Eine Fallstudie*. Zürich: Universität Zürich. Dissertation.
- Assmann, Jan.** 1988. Kulturelles Gedächtnis und kulturelle Identität (engl. version). In Jan Assmann & Tonio Hölscher (eds.). *Kultur und Gedächtnis*, 1. Aufl. (Suhrkamp-Taschenbuch Wissenschaft 724), 9–19. Frankfurt am Main: Suhrkamp. [http://www.fritz-bauer-institut.de/gastprofessur/weissberg/06\\_Jan-Assmann.pdf](http://www.fritz-bauer-institut.de/gastprofessur/weissberg/06_Jan-Assmann.pdf) (accessed December 30th 2009).
- Assmann, Jan.** 1992. *Das kulturelle Gedächtnis: Schrift, Erinnerung und politische Identität in frühen Hochkulturen*, 5. Aufl. dieser Ausg. (Beck'sche Reihe 1307). München: Beck.
- Assmann, Jan.** 2008. Communicative and Cultural Memory. In Astrid Erll & Ansgar Nünning (eds.). *Cultural memory studies. An international and interdisciplinary handbook* (Media and cultural memory 8), 109–118. Berlin: de Gruyter.
- Assmann, Jan and Tonio Hölscher** (eds.). 1988. *Kultur und Gedächtnis*, 1. Aufl. (Suhrkamp-Taschenbuch Wissenschaft 724). Frankfurt am Main: Suhrkamp.
- Baldry, Anthony and Paul J. Thibault.** 2006. *Multimodal transcription and text analysis: A multi-media toolkit and coursebook*. London: Equinox Publ.
- Barthes, Roland.** 1973. *S/Z*. New York: Hill and Wang.
- Barthes, Roland.** 2001. *Image, music, text*, 22. print. New York: Hill and Wang.
- Bazzanella, Carla.** 2002. The significance of context in comprehension. The 'we' case. *Foundations of Science (SI: The context of context)* 7. 239–254.
- Beaman, Karen.** 1984. Coordination and subordination revisited: syntactic complexity in spoken and written narrative discourse. In Deborah Tannen (ed.). *Coherence in spoken and written discourse* (Advances in discourse processes v. 12), 45–80. Norwood, N.J: ALEX Pub. Corp..

- Beißwenger, Michael.** 2007. *Sprachhandlungskoordination in der Chat-Kommunikation*. Univ., Dissertation, Dortmund, 2007. (Linguistik - Impulse & Tendenzen 26). Berlin: de Gruyter.
- Bhatia, Vijay K.** 1993. *Analysing genre: Language use in professional settings*, 1. publ. (Applied linguistics and language study). London: Longman.
- Biber, Douglas.** 1988. *Variation across speech and writing*, 1. paperback ed. Cambridge: Cambridge Univ. Press.
- Biber, Douglas.** 1992. Using computer-based text corpora to analyse the referential strategies of spoken and written texts. In Jan Svartvik (ed.). *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4.-8. August 1991* (Trends in linguistics. Studies and monographs 65), 213–253. Berlin: Mouton de Gruyter.
- Biber, Douglas.** 2006. *Dimensions of register variation: A cross-linguistic comparison*, 1. pb. version. Cambridge: Cambridge Univ. Press.
- Blood, Rebecca.** 2003. *We've got blog. How weblogs are changing our culture*. New York: Perseus Books.
- Bolter, Jay D.** 2001. *Writing space: Computers, hypertext, and the remediation of print*, 2. ed. Mahwah, NJ: Erlbaum.
- Bolter, Jay D. and Richard Grusin.** 2003. *Remediation: Understanding new media*, 6. Nachdr. Cambridge, Mass.: MIT Press.
- Brown, Gillian and George Yule.** 2001. *Discourse analysis*, Repr. (Cambridge textbooks in linguistics). Cambridge: Cambridge Univ. Press.
- Bublitz, Wolfram.** 1994. In the eye of the beholder: the rather mystical notion of coherence. In K. Davidse K. and B. Rudzka-Ostyn Carlon (ed.). *Perspectives on English*, 213–230. Leuven: Peeters.
- Bublitz, Wolfram.** 2000. Cohesion and coherence. *Handbook of Pragmatics* (4). 1-15.
- Bublitz, Wolfram.** 2006. It utterly boggles the mind: knowledge, common ground and coherence. In Hannah Pishwa (ed.). *Language and memory: aspects of knowledge representations. (Trends in Linguistics, Studies and Monographs)*, 359–386. Berlin: de Gruyter.
- Bublitz, Wolfram.** 2008. Sailing the islands or watching from the dock: The treacherous simplicity of a metaphor. How we handle 'new (electronic) hypertext' versus 'old (printed) text'. In Andrea Gerbig & Michael Stubbs (eds.). *Language, people, numbers. Corpus linguistics and society; [on the occasion of Michael W. Stubbs' 60th birthday]* (Language and computers 64). Amsterdam: Rodopi.

- Bucher, Hans-Jürgen.** 1998. Die Zeitung als Hypertext - Verstehensprobleme und Gestaltungsprinzipien für Online-Zeitungen. In Henning Lobin (ed.). *Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering*. Opladen: Westdt. Verl.
- Bühler, Karl.** 1999/1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*, 3. Aufl., ungekürzter Neudr. d. Ausg. Jena, Fischer, 1934. (UTB für Wissenschaft Uni-Taschenbücher 1159). Stuttgart: Lucius & Lucius.
- Bühler, Karl.** 1982. The deictic field of language and deictic words. *Language*, 9-30.
- Burg, Thomas N.** (ed.). 2005. *BlogTalk 2.0: The European conference on weblogs*. Vienna: Cultural Research.
- Bush, Vannevar.** 1945. As we may think. *The Atlantic Monthly* 1 (1-2).
- Carlson, Keith et al** (ed.). 1994. *Perspectives on English*. Leuven: Peeters.
- Cavell, Richard.** 2003. *McLuhan in space: A cultural geography*, Reprinted in paperback with corr. Toronto: Univ. of Toronto Press.
- Chafe, Wallace L.** 1982. Integration and involvement in speaking, writing, and oral literature. In Deborah Tannen (ed.). *Spoken and written language. Exploring orality and literacy* (Advances in discourse processes 9), 35–54. Norwood, NJ: Ablex Publ. Corp.
- Coe, Richard et al** (ed.). 2002. *The rhetoric and ideology of genre. Strategies for stability and change*. Cresskill, NY: Hampton.
- Coiro, Julie et al** (ed.). 2009. *Handbook of research on new literacies*, Repr. New York, NY: Routledge.
- Coulthard, Malcolm** (ed.). 1994. *Advances in written text analysis*. London: Routledge.
- Crowston, Kevin and Marie Williams.** 2000. Reproduced and emergent genres of communication on the world wide web. *The Information Society* 16 (3). 201-,215.
- Cruse, Allan.** 1986. *Lexical semantics* (Cambridge textbooks in linguistics). Cambridge: Univ. Press.
- Crystal, David.** 2001. *Language and the Internet*, 2. ed. Cambridge: Cambridge Univ. Press.
- Crystal, David.** 2005. *The scope of Internet linguistics*. American Association for the Advancement of Science meeting, (18 February 2005) shortened version in *The Western Mail*, February. online version:  
[http://www.davidcrystal.com/DC\\_articles/Internet2.pdf](http://www.davidcrystal.com/DC_articles/Internet2.pdf)
- Deleuze, Gilles and Félix Guattari.** 2007. *A thousand plateaus: Capitalism and schizophrenia*, 12. Minneapolis, Minn.: Univ. of Minnesota Press.

- De Moor, Aldo and Lilian Efimova.** 2004. An Argumentation Analysis of Weblog Conversations. Online publication:  
<https://doc.telin.nl/dscgi/ds.py/Get/File-40067> (accessed December 30th 2009).
- Dillon, Andrew and Barbara Gushrowski.** 2000. Genre and the Web: is the personal home page the first uniquely digital genre? *Journal of The American Society for Information Science* 51 (2). 202-205.
- Dorgeloh, Heidrun and Anja Wanner** (ed.). forthcoming. *Approaches to Syntactic Variation and Genre*. Berlin: Mouton de Gruyter.
- Drew, Paul and Anthony Wootton** (ed.). 1988. *Erwin Goffman. Exploring the Interaction Order*. Boston: Northeastern University Press.
- Drieman, Greg H.** 1962. Differences between written and spoken language. *Acta Psychologica* 20. 78-100.
- Eckert, Penelope.** 2000. *Linguistic variation as social practice: The linguistic construction of identity in Belten High* (Language in Society 27). Malden, Mass.: Blackwell Publishers.
- Eco, Umberto.** 2009. *From Internet to Gutenberg*, The Italian Academy for Advanced Studies in America. New York: Columbia University.  
<http://www.hf.ntnu.no/anv/Finnbo/tekster/Eco/internt2.htm>  
 (accessed December 30th 2009).
- Eisenlauer, Volker and Christian Hoffmann.** 2008. The metapragmatics of remediated text design. *Information Design Journal* 16 (1). 1-18.
- Erickson, Thomas.** 2000. Making sense of computer-mediated communication (CMC): conversations as genres, CMC systems as genre ecologies. *Proceedings of the 33th Hawai'i International Conference on System Sciences (HICSS-33)*. online publication:  
[http://www.pliant.org/personal/Tom\\_Erickson/genreEcologies.html](http://www.pliant.org/personal/Tom_Erickson/genreEcologies.html)  
 (accessed December 30th 2009).
- Erll, Astrid and Ansgar Nünning** (eds.). 2008. *Cultural memory studies: An international and interdisciplinary handbook* (Media and cultural memory 8). Berlin: de Gruyter.
- Fillmore, Charles J.** 1987. Scenes-and-Frames Semantics. In Charles J Fillmore, René Dirven & Günter Radden (eds.). *Fillmore's case grammar. A reader* (Studies in descriptive linguistics 16), Heidelberg: Groos. 79-87.
- Fillmore, Charles J., René Dirven and Günter Radden** (eds.). 1987. *Fillmore's case grammar. A reader* (Studies in descriptive linguistics 16). Heidelberg: Groos.
- Firth, John R.** 1957. Modes of Meaning. *Papers in Linguistics* 1935 -1951. 190 - 215.

- Flood, James** (ed.). 1984. *Understanding reading comprehension: Cognition, language, and the structure of prose*. Newark, Del.: International Reading Association.
- Foltz, Peter**. 1996. Comprehension, Coherence and Strategies in Hypertext and Linear Text. In Jean-François Rouet (ed.). *Hypertext and cognition*, 109 - 137. Mahwah, NJ: Lawrence Erlbaum.
- Freisler, Stefan**. 1994. Hypertext - Eine Begriffsbestimmung. *Deutsche Sprache* 94(1). 19-50.
- Gerbig, Andrea and Michael W. Stubbs** (eds.). 2008. *Language, people, numbers: Corpus linguistics and society; [on the occasion of Michael W. Stubbs' 60th birthday]* (Language and computers 64). Amsterdam: Rodopi.
- Gill, Francis**. 1994. Labelling Discourse: An aspect of nominal-group lexical cohesion. In Malcolm Coulthard (ed.). *Advances in written text analysis*, 83–101. London: Routledge.
- Giltrow, Janet**. 2002. Meta-Genre. In Richard et al Coe (ed.). *The Rhetoric and Ideology of Genre. Strategies for Stability and Change*, 187–295. Cresskill, NY: Hampton.
- Giltrow, Janet and Dieter Stein** (eds.). 2009. *Genres in the Internet. Issues in the theory of genre* (Pragmatics & beyond N.S., 188). Amsterdam: Benjamins.
- Givón, Talmy**. 1990. *Syntax: a functional-typological introduction*. Amsterdam: Benjamins.
- Goethe (von), Johann Wolfgang**. 2009/1832. *Faust II: In fünf Akten*. Hamburg: Hamburger Lesehefte Verlag.
- Goethe (von), Johann Wolfgang**. 2006/1786, *Italienische Reise*. Boston: Adamant Media Corporation.
- Goffman, Erving**. 1982. *Interaction ritual: Essays on face-to-face behavior*. New York: Pantheon Books.
- Gurak, Laura et al** (ed.). 2004. *Into the blogosphere: rhetoric, community, and culture of weblogs*.  
[http://blog.lib.umn.edu/blogosphere/blogging\\_as\\_social\\_action\\_a\\_genre\\_analysis\\_of\\_the\\_weblog.html](http://blog.lib.umn.edu/blogosphere/blogging_as_social_action_a_genre_analysis_of_the_weblog.html) (accessed December 30th 2009).
- Gutwinski, Waldemar**. 1976. *Cohesion in Literary Texts*. The Hague: Mouton.
- Halbwachs, Maurice**. 2006. *On collective memory*, [reprint] (The Heritage of sociology). Chicago: University of Chicago Press.
- Halliday, Michael A. K. and Ruqaiya Hasan**. 1976. *Cohesion in English*. London: Longman.

- Hasan, Ruqaiya.** 1984. Coherence and cohesive harmony. In James Flood (ed.). *Understanding reading comprehension. Cognition, language, and the structure of prose*, 181–219. Newark, Del.: International Reading Association.
- Haugeland, John** (ed.). 1988. *Mind design: Philosophy, psychology, artificial*, 6. print. (A Bradford book). Cambridge (Mass): MIT Press.
- Herman, David** (ed.). 2005. *Routledge encyclopedia of narrative theory*, 1. publ. in paperback. London: Routledge.
- Herring, Susan et al** 2004. Bridging the gap: A genre analysis of weblogs. *Proceedings of the 37th Hawai'i International Conference on System Sciences (HICSS-37)*(37). <http://www.blogninja.com/DDGDD04.doc> (accessed December 30th 2009).
- Herring, Susan et al** 2005. Conversations in the Blogosphere: An Analysis "From the Bottom Up". *Proceedings of the Thirty-Eighth Hawai'i International Conference on System Sciences (HICSS-38)*(38). online publication. <http://ella.slis.indiana.edu/~herring/blogconv.pdf> (accessed December 30th 2009).
- Herring, Susan and John Paolillo.** 2006. Gender and genre variation in weblogs. *Journal of Sociolinguistics* 10 (4), 439-459.
- Heydt, Theresa.** 2009. A model for describing 'new' and 'old' properties of CMC genres: the case of digital folklore. In Janet Giltrow & Dieter Stein (eds.). *Genres in the Internet. Issues in the theory of genre* (Pragmatics & beyond N.S., 188), 239-263. Amsterdam: Benjamins.
- Hiraga, Masako.** 1994. Diagrams and metaphors: Iconic aspects in language. *Journal of Pragmatics* 24. 5-21.
- Hoey, Michael.** 2005. *Textual interaction: An introduction to written discourse analysis*. London: Routledge.
- Holzner, Johann** (ed.). *Studien zur Literatur des 19. und 20. Jahrhunderts in Österreich: Festschrift für Alfred Doppler zum 60. Geburtstag*. Innsbruck: Institut für Germanistik.
- Howard, Marc W. et al.** 2005. *Semantic Structure and Episodic Memory*. Online publication: <http://memory.syr.edu/marc/papers/HowaAddiJingKaha-LSAChap-doc.pdf> (accessed December 30th 2009).
- Huber, Oliver.** 2002. *Hyper-Text-Linguistik: TAH: ein textlinguistisches Analysemodell für Hypertexte. Theoretisch und praktisch exemplifiziert am Problemfeld der typisierten Links von Hypertexten im World Wide Web. Dissertation*. Munich: LMU Munich Dissertation. <http://edoc.ub.uni-muenchen.de/921/> (accessed December 30th 2009).
- Kennedy, Graeme.** 1998. *An introduction to corpus linguistics* (Studies in language and linguistics). London: Longman.

- Kittler, Friedrich.** 1996. The history of communication media. *cttheory.net* (114). no pagination. <http://www.cttheory.net/articles.aspx?id=45> (accessed December 30th 2009).
- Koch, Peter and Wulf Oestereicher.** Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgebrauch. *Romanistisches Jahrbuch* 36. 15-43.
- Kolb, David.** 1994. *Socrates in the labyrinth: Hypertext, Argument, Philosophy*. Watertown, Mass.: Eastgate Systems.
- Kuhlen, Rainer.** 1991. *Hypertext. Ein nichtlineares Medium zwischen Buch und Wissen- schaft*. Berlin: Springer.
- Landow, George P.** 1997. *Hypertext 2.0: Hypertext - the convergence of contemporary critical theory and technology*, Rev., amplified ed. (Parallax). Baltimore, Md.: Johns Hopkins Univ. Press.
- Lange, Patricia.** 2007. Publicly private and privately public: social networking on youtube. *Journal of Computer-Mediated Communication* 13(1). online publication: <http://jcmc.indiana.edu/vol13/issue1/lange.html> (accessed December 30th 2009).
- Leech, Geoffrey.** 1981. *Semantics: The study of meaning*, 2nd ed., rev. and updated. Harmondsworth: Penguin Books.
- Lemke, Jay.** 2002. Travels in Hypermodality. *Visual Communication* 1(3). 299-325.
- Lemke, Jay.** 2005. Multimodal Genres and Traversals. In Eija Ventola et al (ed.). *Approaches to Genre (Folia Linguistica)* (39), 45-56.
- Levinson, Stephen C.** 1988. Putting linguistics on a proper footing. Explorations in Goffman's concepts of participation. In: Paul and Wootton Drew (ed.). *Erwin Goffman. Exploring the Interaction Order*, 161–227. Boston: Northeastern University Press.
- Lipka, Leonhard.** 2002. *English lexicology: Lexical structure, word semantics & word-formation*, 3. ed. (Narr-Studienbücher). Tübingen: Narr.
- Lobin, Henning** (ed.). 1998. *Text im digitalen Medium: Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering*. Opladen: Westdt.
- Löbner, Sebastian.** 2007. *Understanding semantics*, [reprint] (Understanding language series). London: Hodder Education.
- Lodge, David.** 1981. *Changing places: A tale of 2 campuses*, Reprint. Harmondsworth: Penguin.
- Lyons, John.** 1977. *Semantics*. Cambridge: Cambridge Univ. Press.



- Martin, James R. and Gordon Williams.** 2004. Functional sociolinguistics. In Ulrich Ammon et al (ed.). *Sociolinguistics. An international handbook of the science of language and society*, 120-129. Berlin: de Gruyter.
- Martin, James R.** 1992. *English text: System and structure*. Philadelphia: Benjamins.
- Martin, James R.** 2001. Cohesion and texture. In Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton (eds.). *The handbook of discourse analysis* (Blackwell handbooks in linguistics), Malden, Mass.: Blackwell. 35-54.
- Matsui, Tomoko.** 1993. Bridging, reference and the notions of topic and focus. *Lingua* 9(1(2)). 49-68.
- Matsui, Tomoko.** 2000. *Bridging and relevance*. 1995 (Diss.), 1995. (Pragmatics & beyond new ser. 76). Amsterdam, Philadelphia: Benjamins.
- McLuhan, Marshall.** 1994/1964. Understanding media. The extensions of men. Cambridge (MA): MIT Press.
- McNeil, Laurie.** 2009. Brave new genre or generic colonialism? Debates over ancestry in Internet Diaries. In Janet Giltrow & Dieter Stein (eds.). *Genres in the internet. Issues in the theory of genre* (Pragmatics & beyond N.S., 188), Amsterdam: Benjamins. 143-163.
- Miller, Carolyn R. and Dawn Sheperd.** 2004. Blogging as social action: a genre analysis of the weblog. In Laura et al Gurak (ed.). *Into the blogosphere: rhetoric, community, and culture of weblogs*, online publication. Online publication: [http://blog.lib.umn.edu/blogosphere/blogging\\_as\\_social\\_action\\_a\\_genre\\_analysis\\_of\\_the\\_weblog.html](http://blog.lib.umn.edu/blogosphere/blogging_as_social_action_a_genre_analysis_of_the_weblog.html) (accessed December 30th 2009).
- Miller, Carolyn R. and Dawn Sheperd.** 2009. Questions for genre theory from the blogosphere. In Janet Giltrow & Dieter Stein (eds.). *Genres in the Internet. Issues in the theory of genre* (Pragmatics & beyond N.S., 188), Amsterdam: Benjamins. 263-291.
- Minsky, Marvin.** 1988. A framework for representing knowledge. In John Haugeland (ed.). *Mind design. Philosophy, psychology, artificial*, 6. print. (A Bradford book), Cambridge, Mass.: MIT Press. 111–143.
- Mortensen, Torill.** 2009. Of a divided mind: weblog literacy. In Julie Knobel M. and Lankshear C. Coiro (ed.). *Handbook of research on new literacies*, Repr. New York, NY: Routledge.
- Nelson, Theodor H.** 1965. A file structure for the complex, the changing and the indeterminate. *Proceedings of the 20th National ACM-Conference* (20). 84 -100.
- Nelson, Theodor H.** 1987. *Literary machines*, Ed. 87.1. South Bend, Sausalito (CA): Mindful Press.
- Nilsson, Stephanie.** 2003. *The function of language to facilitate and maintain social networks in research weblogs*. D-Essay. University of Umea.

- O'Reilly, Tim.** 2005. *What Is Web 2.0?: Design patterns and business models for the next generation of software*. Online publication:  
<http://oreilly.com/web2/archive/what-is-web-20.html> (accessed December 30th 2009).
- Ong, Walter J.** 2003/1988. *Orality and literacy: The technologizing of the word*, Reprinted. (New accents). London: Routledge.
- Painter, Clare and James Martin.** 1986. Writing to mean: Teaching Genres across the Curriculum. *Applied Linguistics Association of Australia. Occasional Papers* (9).
- Pishwa, Hannah** (ed.). 2006. *Language and Memory: Aspects of Knowledge Representations. (Trends in Linguistics, Studies and Monographs)*. Berlin: de Gruyter.
- Puschmann, Cornelius.** Forthcoming. "Thank you for thinking we could". Use and function of interpersonal pronouns in corporate weblogs. In Heidrun Dorgeloh and Anja Wanner (eds). *Approaches to Syntactic Variation and Genre*. Berlin: Mouton de Gruyter.
- Puschmann, Cornelius.** Forthcoming (b). *The corporate blog as an emerging genre of computer-mediated communication: features, constraints, discourse situation*. (Göttinger Schriften zur Internetforschung). Göttingen: Universitätsverlag Göttingen.
- Puschmann, Cornelius.** 2009. Lies at Walmart. Style and the subversion of genre in the Life at Wal-Mart blog. In Janet Giltrow & Dieter Stein (eds.). *Genres in the internet. Issues in the theory of genre* (Pragmatics & beyond N.S., 188), 49–85. Amsterdam: Benjamins.
- Rosch, Eleanor.** 1978. *Cognition and categorization*. Hillsdale, NJ: Erlbaum [u.a.].
- Rouet, Jean-François** (ed.). 1996. *Hypertext and cognition*. Mahwah, NJ: Lawrence Erlbaum.
- Schiffrin, Deborah, Deborah Tannen and Heidi E. Hamilton** (eds.). 2001. *The handbook of discourse analysis* (Blackwell handbooks in linguistics). Malden, Mass.: Blackwell.
- Schmid, Hans-Jörg.** 1999. Cognitive effects of shell nouns. In Karen VanHoek, Andrej A Kibrik & Leonard G. M Noordman (eds.). *Discourse studies in cognitive linguistics. Selected papers from the Fifth International Cognitive Linguistics Conference, Amsterdam, July 1997* (Amsterdam studies in the theory and history of linguistic science Series 4, Current issues in linguistic theory 176), 111–132. Amsterdam: Benjamins.
- Schmidt, Jan.** 2006. *Weblogs. Eine kommunikationssoziologische Studie*. Konstanz: UVK.

- Schubert, Christoph.** 2008. *Englische Textlinguistik: Eine Einführung* (Grundlagen der Anglistik und Amerikanistik 30). Berlin: Schmidt.
- Sinclair, John M.** 1991. *Corpus, concordance, collocation* (Describing English language). Oxford: Oxford Univ. Press.
- Smith, Raoul and William J. Frawley.** 1983. Conjunctive cohesion in four English genres. *Text* (3). 347-374.
- Soanes, Catherine and Angus Stevenson** (eds.). 2003. *The Oxford Dictionary of English*. Oxford: Oxford University Press.
- Stein, Dieter.** 2006. The website as a domain-specific Genre. *Language@Internet* 3(6). no pagination. online publication:  
<http://www.languageatinternet.de/articles/2006> (accessed December 30th 2009).
- Storrer, Angelika.** 2002. Coherence in text and hypertext. *Document Design* 3(2). 156-168.
- Stubbs, Michael.** 1998. *Discourse analysis: The sociolinguistic analysis of natural language*, Reprinted. (Language in society 4). Oxford: Basil Blackwell.
- Stubbs, Michael.** 2001. *Words and phrases: corpus studies of lexical semantics*. Oxford u.a: Blackwell.
- Svartvik, Jan** (ed.). 1992. *Directions in corpus linguistics: proceedings of the Nobel Symposium 82, Stockholm, 4.-8. August 1991* (Trends in linguistics, Studies and monographs 65). Berlin: Mouton de Gruyter.
- Swales, John M.** 1990. *Genre analysis: English in academic and research settings*, 1. publ. (The Cambridge applied linguistics series). Cambridge: Cambridge Univ. Press.
- Tannen, Deborah** (ed.). 1982. *Spoken and written language: Exploring orality and literacy* (*Advances in discourse processes* 9). Norwood, New Jersey: Ablex Publ. Corp..
- Tannen, Deborah** (ed.). 1984. *Coherence in spoken and written discourse* (*Advances in discourse processes* 12). Norwood, New Jersey: ABLEX Pub. Corp..
- Tanskanen, Sanna-Kaisa.** 2006. *Collaborating towards coherence*. Amsterdam: John Benjamins.
- Tottie, Gunnel and Ingegerd Bäcklund** (eds.). 1986. *English in speech and writing: a symposium*. Stockholm: Almquist and Wiksell.
- Trammell, Kaye D. and Ana Keshelashvili.** 2005. Examining the new influencers: a self-presenta- tion study of A-list blogs. *Journalism and Mass Communication Quarterly* (82) 4. 968-982.

- Tulving, Endel.** 1983. *Elements of episodic memory* (Oxford psychology series 2). Oxford: Clarendon Press.
- Tulving, Endel.** 2002. Episodic memory: from mind to brain. *Annual Review of Psychology*(53). 1–25.
- Ungerer, Friedrich and Hans-Jörg Schmid.** 2007. *An introduction to cognitive linguistics*, 2. ed., [reprint] (Learning about language). Harlow, England: Pearson Longman.
- van Leeuwen, Theo.** 2008. *Discourse and practice: New tools for critical discourse analysis* (Oxford studies in sociolinguistics). Oxford: Oxford Univ. Press.
- VanHoek, Karen, Andrej A. Kibrik and Leonard G. M. Noordman** (eds.). 1999. *Discourse studies in cognitive linguistics: Selected papers from the Fifth International Cognitive Linguistics Conference, Amsterdam, July 1997* (Amsterdam studies in the theory and history of linguistic science, Series 4, Current issues in linguistic theory 176). Amsterdam: Benjamins.
- Ventola, Eija.** 1987. *The structure of social interaction: A systemic approach to the semiotics of service encounters*, 1. publ. London: Pinter.
- Ventola, Eija et al** (ed.). 2005. *Approaches to genre (Folia Linguistica)* (39).
- Voiskounsky, Alexander.** 1997. Telelogue conversations. *Journal of Computer-Mediated Communication* 2(4). no pagination (Online Journal).  
<http://jcmc.indiana.edu/vol2/issue4/voiskounsky1.html> (accessed December 30th 2009).
- Walker, Jill.** 2005. Weblog. In David Herman (ed.). *Routledge encyclopedia of narrative theory*. London: Routledge.
- Walker-Rettberg, Jill.** 2008. *Blogging. Digital Media and Society*. Cambridge: Polity Press.
- Widdowson, Henry G.** 2008. *Discourse analysis*, (Oxford introductions to language study). Oxford: Oxford Univ. Press.
- Wijnia, Elmine.** 2005. Understanding Weblogs: A Communicative Perspective. In Thomas N. Burg (ed.). *BlogTalk 2.0. The European Conference on Weblogs*, 38–83. Vienna: Cultural Research.
- Wittgenstein, Ludwig and Gertrude E. M. Anscombe.** 2004/1953. *Philosophical investigations: The German text, with a rev. English translation*, 3. ed., 50th anniversary commemorative ed., repr. Malden, MA: Blackwell.
- Wolf, Norbert R.** 1993. Am Beispiel Elias Canettis: Überlegungen zur Textsyntax und Texttypologie. In Johann Holzner (ed.). *Studien zur Literatur des 19. und 20. Jahrhunderts in Österreich: Festschrift für Alfred Doppler zum 60. Geburtstag.*, 205-218. Innsbruck: Institut für Germanistik.

## Webliography

---

### Weblogs (AWC)

Ohmyword (OMW): <http://omywordblog.blogspot.com/> (December 30<sup>th</sup> 2009)  
 Shelli's Sentiments (ShSe): <http://shellis-sentiments.com/> (December 30<sup>th</sup> 2009)  
 Beetle's Memories and Ravings (BMR): <http://beetle-blog.com/> (December 30<sup>th</sup> 2009)  
 Mushy's Moochings (MM): <http://www.mushysmoochings.blogspot.com/> (December 30<sup>th</sup> 2009)  
 All Adither (AA): <http://alladither.typepad.com/> (December 30<sup>th</sup> 2009)  
 NonSoccerMom (NSM): [www.nonsoccermum.com](http://www.nonsoccermum.com) (December 30<sup>th</sup> 2009)  
 WhateverIThink (WIT): <http://joanharvest.wordpress.com/> (December 30<sup>th</sup> 2009)  
 FiveBlondes (FB): <http://fiveblondes.com/> (December 30<sup>th</sup> 2009)  
 BlogofaGoodTime (BOGT): [http://julie\\_gong.blogspot.com/](http://julie_gong.blogspot.com/) (December 30<sup>th</sup> 2009)  
 BackHomeAgain (BHA): <http://jwiley.typepad.com> (December 30<sup>th</sup> 2009)

### Additional Weblogs Used for Illustrative Purposes

<http://jilltxt.net/> (December 30<sup>th</sup> 2009)  
<http://ifitshipitshere.Weblogspot.com/> (December 30<sup>th</sup> 2009)  
<http://kariandkijisa.Weblogspot.com/> (December 30<sup>th</sup> 2009)  
<http://oneminutewriter.Weblogspot.com/> (December 30<sup>th</sup> 2009)  
<http://doctor-chef.Weblogspot.com> (December 30<sup>th</sup> 2009)  
<http://capecodmemories.Weblogspot.com/> (December 30<sup>th</sup> 2009)  
<http://theglassdoorknob.Weblogspot.com> (December 30<sup>th</sup> 2009)

**Appendix**

---

- I. Online Questionnaire (In Print)
- II. Result Sheets: Grammatical Cohesion (CD-Rom)
- III. Results Sheets: Lexical Cohesion (CD-Rom)

The CD-Rom contains:

- (a) Full Version of the PhD thesis (.doc/.docx/.pdf)
- (b) Result Files → Grammatical Cohesion (Entries and Comments)
- (c) Result Files → Lexical Cohesion (Entries and Comments)

## Digital Questionnaire

(distributed through surveymonkey.com)

---

**\* 1. What is the name of your weblog?**

**2. What's your age? (\*not mandatory)**

- ☐ 20-30
- ☐ 30-40
- ☐ 40-50
- ☐ 60-70

**3. How long have you been blogging?**

- ☐ 1-2 years
- ☐ 3-4 years
- ☐ 5-6 years
- ☐ > 6 years

**4. Why do you keep a weblog? Please rank!**

	1 (not at all)	2	3	4	5	6	7	8	9	10 (absolutely)
to express yourself creatively	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to document your pers. experiences & share them with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to stay in touch with friends and family	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to share practical skills and knowledge with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to motivate other people to action	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to entertain people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to store resources or information that is important to you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to influence the way other people think	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to network and meet new people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
to make money	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**5. How tech-savvy would you consider yourself?**

	1 (beginner)	2	3	4	5 advanced	6	7	8	9	10 (professional)
Web-related tech-knowledge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wish to specify?

**6. If you had to decide. Would you see your weblog mainly as a...**

- ☐ private diary
- ☐ public journal
- ☐ friendship network
- ☐ communication hub
- ☐ collection of topic-related hyperlinks
- ☐ a diary and a link collection

**7. Why do you think you use pictures/photographs in your blog?**

- ☐ to boost their visual appeal
- ☐ to illustrate verbal descriptions
- ☐ to document personal experiences
- ☐ to humour other users
- ☐ to show who you are as author
- ☐ to create continuity between posts

Other (please specify)

**8. When using pictures in your blog: which of the following criteria seem most relevant to you?**

- ☐ overt relation between image and surrounding verbal text
- ☐ placement in posts (top, middle, bottom of post)
- ☐ size of pictures (large, medium, small)
- ☐ picture resolution
- ☐ picture content (objects, persons, etc.)
- ☐ framing (perspective, size, lighting)
- ☐ body language (gestures and mimics)
- ☐ composition of pictorial elements (vectors, object relations of pictorial elements)

Other (please specify)



**9. Do you know all of the users who comment in blog?**

- ☐ I know them all
- ☐ I know most of them
- ☐ I know only few
- ☐ I don't know them at all

Are they internet acquaintences or real-life friends?

**10. Do you have different target readerships in mind when you're composing a post? (e.g. familiy, friends, colleagues, internet acquaintences, etc.)**

- ☐ Always
- ☐ Sometimes
- ☐ Never

Specifications?



